

THE ACCEPTANCE AND CONTINUED USAGE OF OPEN SOURCE
DATA SCIENCE AND ADVANCED ANALYTICS TECHNOLOGIES

RIZA ERGÜN ARSAL

BOĞAZIÇI UNIVERSITY

2022

THE ACCEPTANCE AND CONTINUED USAGE OF OPEN SOURCE
DATA SCIENCE AND ADVANCED ANALYTICS TECHNOLOGIES

Thesis submitted to the
Institute for Graduate Studies in Social Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Management

by
Riza Ergün Aarsal

Boğaziçi University

2022

DECLARATION OF ORIGINALITY

I, Rıza Ergün Arsal, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date

ABSTRACT

The Acceptance and Continued Usage of Open Source Data Science and Advanced Analytics Technologies

Due to the ubiquity of open source technologies in data science, organizations started to lean towards open source ecosystems. With the increasing number of citizen data scientists in organizations, individuals in business units with non-technical backgrounds also need to adopt and utilize these technologies. This signifies a substantial change for individuals who are used to working with proprietary technologies for which relying on a major vendor is inherently more reassuring. To address this problem, we develop and test a model that examines the trust-based determinants of the continued use of open source data science and advanced analytics technologies (OSDST) in organizations. We conducted a field survey methodology for data collection purposes. We used Partial Least Squares Structured Equation Modeling (PLS-SEM) to test the model's validity and the hypothesized relationships. Our results indicate that trust in OSDST strongly influences usefulness perceptions and user satisfaction. We have also found that situational normality and trust in the data science community are two direct determinants of trust in OSDST. Further, compared to the top management championship, we found that the championship of co-workers plays a more central role in shaping trusting beliefs. To the best of our knowledge, this is the first study investigating the individual use of OSDST in organizations, where most individuals actively adopt and utilize open source technologies for the first time.

ÖZET

Açık Kaynak Kodlu Veri Bilimi ve İleri Analitik

Teknolojilerinin Benimsenmesi ve Kullanılması

Veri biliminde açık kaynak kodlu teknolojilerin yaygınlaşmasıyla beraber kurumlar açık kaynak ekosistemlerine yönelmeye başladı. Organizasyonlardaki vatandaş veri bilimcilerinin sayısının artmasıyla birlikte iş birimlerinde yeterli teknik altyapıya sahip olmayan bireylerin de bu açık kaynak kodlu teknolojileri benimseme ve kullanma ihtiyacı doğdu. Bu durum büyük tedarikçi firmalar tarafından sağlanan ve bu yüzden de doğası gereğiyle daha çok güven veren teknolojilerle çalışmaya alışmış bireyler için önemli bir değişiklik ifade etmektedir. Bu sorunu ele almak için kurumlardaki açık kaynak kodlu veri bilimi ve ileri analitik teknolojilerin (OSDST) kullanımını etkileyen güvene dayalı faktörleri araştıran bir model geliştirdik ve test ettik. Veri toplamak için saha araştırması yürüttükten sonra modelin geçerliliğini ve hipotezlerini test etmek için Kısmi En Küçük Kareler-Yapısal Eşitlik Modellemesi (PLS-SEM) yöntemini kullandık. Sonuçlar, kullanılan açık kaynak kodlu teknolojilere duyulan güvenin bu teknolojilerin hem algılanan faydaları hem de kullanıcı memnuniyeti üzerinde güçlü bir etkisi olduğunu göstermektedir. Bulgular, durumsal normalliğin ve veri bilimi komünitelerine duyulan güvenin, halihazırda kullanılmakta olan açık kaynak kodlu teknolojilere duyulan güveni pozitif olarak etkilediğine işaret etmektedir. Bu araştırma, bilgimiz dahilinde, açık kaynak kodlu veri bilimi ve ileri analitik teknolojilerinin kurumlardaki bireysel kullanımını araştıran ilk çalışmadır.

CURRICULUM VITAE

NAME: Rıza Ergün Aرسال

DEGREES AWARDED

PhD in Management, 2022, Boğaziçi University

MS in Information Systems Management, 2002, Ferris State University

BS in International Relations, 1998, Istanbul University

AREAS OF SPECIAL INTEREST

Data science, advanced analytics, adoption and utilization of IT in organizations, online consumer behavior, trust

PROFESSIONAL EXPERIENCE

Partner and Head of Strategy and Organizational Development, Accelera, 2016-

Instructor and Education Specialist, Istanbul Bilgi University 2010-

Graduate Research and Teaching Assistantship, Department of Management, Clemson University, 2003-2008

Graduate Assistantship, Ferris State University, 2000-2002

AWARDS AND HONORS

Selected to Who's Who Among Students in American Universities & Colleges, 2002

Recipient of Graduate School Excellence Award, Ferris State University, 2002

PUBLICATIONS

Journal Articles

Zimmer, C., Arsal, R.E, AlMarzouq, M., and Grover, V. (2010). "Investigating Online Information Disclosure: Effects of Information Relevance, Trust and Risk". *Information & Management*. 47(2): 115-123

Zimmer, C., Arsal, R.E, AlMarzouq, M., Moore, D., and Grover V. (2010). "Knowing Your Customers: Using a Reciprocal Relationship to Enhance Voluntary Information Disclosure." *Decision Support Systems*. 48(2): 395-406

Thatcher, J.B., McKnight, D.H., Arsal, R.E, Baker, E.W., and Roberts, N. (2009). "The Role of Trust in Post-Adoption IT Exploration: An Empirical Examination of Knowledge Management Systems." *IEEE Transactions on Engineering Management*.

Arsal, R.E, Thatcher, J.B., Zagenczyk, T., McKnight, D.H., and Ahuja, M. (2009). "Organizational Factors and Information Technology Use: Tying Perceptions of the Organization to Perception of IT." *Journal of Organizational and End User Computing*. 21(3): 37-59.

Book Chapters

Arsal, R.E, Thatcher, J.B., Zagenczyk, T., McKnight, D.H., and Ahuja, M. (2011). "Organizational Factors and Information Technology Use: Tying Perceptions of the Organization to Perception of IT." *Book Chapter on Organizational and End-User Interactions: New Explorations*. IGI Global, Chapter 12: 248-271.

Conference Proceedings

Arsal, R.E, Durdu, H.T., and Tongarlak, M.H. (2022). *Organizational Enablers of Data-Driven Digital Transformation: A Case Study from Banking Industry*. IEEE Technology & Engineering Management Conference (TEMSCON), 2022.

Arsal, R.E, Zimmer, C., AlMarzouq, M., and Grover, V. (2007). "An Empirical Investigation of the Determinants of Online Information Disclosure." In the *Proceedings of the Americas Conference on Information Systems*, Keystone, CO.

Arsal, R.E, Thatcher, J.B., and McKnight, D.H. (2004). "Trust in Technology: An Empirical Examination of the Construct." Presented at the Diffusion Interest Group in Information Technology (DIGIT) Workshop, Washington D.C.

ACKNOWLEDGEMENTS

First and foremost, I want to express my gratitude to my advisor Dr. Hayri Tongarлак for his continuous guidance, encouragement, and support throughout the whole journey. His positive and friendly approach helped me deal with the difficulties I faced while completing this study. I cannot thank him enough.

Prof. Dr. Stefan Koch is the first faculty member that I met when I decided to pursue my PhD studies at Boğaziçi University. He was also my former co-advisor until he returned to Austria. I am grateful to him for believing in me and shaping my research ideas with his invaluable feedback.

Prof. Dr. Ayşegül Toker has also been there from the beginning until the completion of this dissertation. I have learned so much from our discussions. Regardless of her several roles such as a dean or a committee member, she always kept her door and phone open whenever I needed.

I also would like to express my thanks to my committee members, Prof. Dr. Şule Işınsu Özmen and Prof. Dr. Fatih Alagöz for agreeing to join my dissertation committee and providing excellent feedback when I was trying to put everything together.

Tolga Durdu deserves a special mention. We started the PhD program together and now we are graduating at the same time. He has been an amazing friend and a true comrade-in-arms in every possible way. We spent countless days and nights working side-by-side on our dissertations. Without him, the whole process would have been much more challenging than it already was.

ACKNOWLEDGEMENTS

I also cannot thank enough to my wonderful family and friends. I am particularly grateful to Hürsu Öke, Emre Erbabacan, Afşin Akın, Okan Kutay, Fehmi Tıǒlı, Tolga Akyıldız, and my brother Kerem Arsal for helping me cross the finish line with their endless support and encouragement.

Thanks to my dear friends Chris Zimmer, Ramakrishna Ayyagari, and Michael Crino for the great times at Clemson University as well as their support after my return to Turkey.

I also want to thank to my cohorts at Bilgi University, Metehan Sekban, Gizem Turna Cebeci, Murat Unanoǒlu, Pınar Yıldırım, and Aslı Elif Aydın Altıoklar for creating such an enjoyable work environment.

Finally, I am indebted to everyone who participated in this study.

DEDICATION

This dissertation is dedicated to:

my wife Şebnem Usluer Arsal, the brightest source of light in my life

and

my parents, who have always been there for me.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Overview.....	1
1.2 Research objectives and questions.....	5
1.3 Proposed research model	5
1.4 Contributions	6
1.5 Outline of dissertation.....	7
CHAPTER 2: LITERATURE REVIEW	8
2.1 Data science and advanced analytics.....	9
2.2 Continued use and business value of IT	15
2.3 IT adoption and implementation in organizations	17
2.4 Organizational level adoption.....	19
2.5 Individual level adoption	22
2.6 Trust in technology	31
2.7 IT championship in organizations	36
2.8 Individual adoption of open source software in IS research.....	38
CHAPTER 3: HYPOTHESES DEVELOPMENT	41
CHAPTER 4: METHODOLOGY	49
4.1 Research design	49
4.2 Sampling strategy	51
4.3 Survey instrument development	54

4.4 Survey administration	66
4.5 Data analysis procedure	67
CHAPTER 5: DATA ANALYSIS AND RESULTS	70
5.1 Data screening	70
5.2 Sample characteristics	71
5.3 Measurement model.....	74
5.4 Structural model.....	83
CHAPTER 6: CONCLUSION.....	91
6.1 Summary of findings	91
6.2 Academic and practical contributions	92
6.3 Limitations and future research	93
APPENDIX A: ETHICS COMMITTEE APPROVAL.....	95
APPENDIX B: SURVEY ADMINISTRATION	96
APPENDIX C: SURVEY INSTRUMENT IN TURKISH.....	97
APPENDIX D: NORMAL DISTRIBUTION MEASURES	101
APPENDIX E: OUTER LOADINGS.....	103
APPENDIX F: DISCRIMINANT VALIDITY: CROSS LOADINGS	105
APPENDIX G: MULTICOLLINEARITY – OUTER VIF VALUES.....	107
REFERENCES.....	110

LIST OF TABLES

Table 1. Top Data Science Tools for 2022	13
Table 2. Summary of IT Implementation Stages	19
Table 3. Selected Technology Acceptance Models	24
Table 4. TAM and UTAUT in Various Contexts	26
Table 5. Trusting Beliefs: Dimensions and Conceptualizations	33
Table 6. Construct Definitions	40
Table 7. List of Hypotheses	48
Table 8. Research Process of the Study	51
Table 9. PLS-SEM vs CB-SEM.....	69
Table 10. Respondents by Age	72
Table 11. Respondents by Gender	72
Table 12. Respondents by Education	73
Table 13. OSDST Used in Organizations	73
Table 14. Respondents by Industry	74
Table 15. Construct Reliability and Convergent Validity.....	75
Table 16. Discriminant Validity: Hetero-Monotrait Ratio (HTMT).....	79
Table 17. Discriminant Validity: Fornell-Larcker Criterion.....	80
Table 18. Reliability and Validity: Summary	81
Table 19. Path Coefficients and Statistical Significance	84
Table 20. Hypotheses Testing Results	86
Table 21. Effect Sizes (f^2)	87
Table 22. Group Differences (Business vs IT Units).....	88

LIST OF FIGURES

Figure 1. Organizational Data Science Models.....	4
Figure 2. Proposed Research Model	6
Figure 3. IS Process Model	15
Figure 4. Updated IS Success Model	16
Figure 5. IT Implementation Model.....	17
Figure 6. DOI Model for IT Adoption	20
Figure 7. TOE Framework	21
Figure 8. Technology Acceptance Model	23
Figure 9. UTAUT.....	25
Figure 10. Feature-Centric View of IT Adoption and Use	27
Figure 11. Expectation-Confirmation Theory.....	29
Figure 12. IS Continuance Model	30
Figure 13. Nomological Network of Trust in Technology	34
Figure 14. Research Model and Hypotheses.....	41

ABBREVIATIONS

AST: Adaptive Structuration Theory

AVE: Average Variance Extracted

CB-SEM: Covariance Based Structured Equation Modeling

DOI: Diffusion of Innovation

ECT: Expectation Confirmation Theory

f²: Effect Size

GEXP: Exploration (General)

CEXP: Exploration (Current)

IS: Information Systems

IT: Information Technology

MCB: Top Management Championship- Belief

MCP: Top Management Championship-Participation

MIS: Management Information Systems

OLS: Ordinary Least Squares

OSDST: Open Source Data Science Technologies

OSS: Open Source Software

PCB: Peer (Co-Worker) Championship- Belief

PCP: Peer (Co-Worker) Championship- Participation

PLS-MGA: Partial Least Squares-Multi-Group Analysis

PLS-SEM: Partial Least Squares-Structured Equation Modeling

PU: Perceived Usefulness

R²: R Squared

SAT: Satisfaction

SEM: Structural Equation Modeling

SIPT: Social Information Processing Theory

TAM: Technology Acceptance Model

TCOM: Trust in Data Science Community

TTP: Trust in Technology-Predictability

TTF: Trust in Technology-Functionality

TTH: Trust in Technology-Helpfulness

TOE: Technology, Organization, Environment

TRA: Theory of Reasoned Action

TSN: Trust-Situational Normality

UTAUT: Unified Theory of Acceptance and Use of Technology

VIF: Variance Inflation Factor

CHAPTER 1

INTRODUCTION

1.1 Overview

Data enables managers to make their decisions based on evidence. Data-driven decision-making involves making relevant data accessible to key decision-makers and providing them with insights about their customers and the business environment. It allows organizations to foster innovation as well as the development of new products and services that would increase customer satisfaction and loyalty (Brynjolfsson, Hitt, & Kim, 2011). Therefore, it has been identified as a critical factor to have the potential to revolutionize management (McAfee & Brynjolfsson, 2012).

The exponentially accelerating advances in computing and telecommunication technologies have paved the way not only for the digitization of information but also the access and handling of ever-growing volumes of data. Thanks to developments in technology, terms such as data science, advanced analytics, and big data have emerged as important business terms that have drawn the interest of scholars and practitioners. Instead of gaining limited insights from small data sets, organizations now have opportunities to develop capabilities which allow them to collect, store, and process massive quantities of data (Bean, 2022). As a result, storing and analyzing a wide array of internally generated and externally collected data have become an invaluable business practice, making data science a ubiquitous phenomenon in academic and business circles over the past decade.

The competitive advantage that can be gained by utilizing all types and sizes of data has resulted in the emergence of data science as a strategic initiative that can

encompass even the entire organizations (Grover, Chiang, Liang, & Zhang, 2018). Statistics provided by one of the world's leading analytics consultancy companies, Booz Allen Hamilton (2017), demonstrate the immense impact of data science practices in organizations: While increasing data access and usability results in an increase of 11% to 49% in productivity and return on assets, firms that can successfully (a) incorporate big data, (b) deploy data analytics across most of the organization, and (c) align daily operations with senior management's goals at the same time enjoy an increase of around 1000% in their return on investments.

Data science and advanced analytics involve the utilization of various technologies for data collection, storage, processing, analysis, and deployment purposes. Due to their advantages such as high quality, security, reliability, flexibility, and no vendor lock-ins (Gwebu & Wang, 2010), most organizations are turning to open source technologies to utilize data. The Apache family that hosted big data projects Spark, Hadoop, Cassandra, Kafka, and Storm emerged as one of the first communities to show that open source is not only a viable alternative but also the future of big data analytics (Augur, 2016). Within the next several years, these assertions have come fully true. Today, scalable technologies for the efficient management and storage of large data sets are available as open source solutions, underpinning the technology stack that powers modern data science (Piatetsky, 2019). Open source platforms such as Hadoop and Spark, development environments such as Jupyter and Zeppelin, and programming languages such as Python and R have become the go-to technologies for many firms who had traditionally invested in proprietary software from IT incumbents such as IBM, Oracle, SAP and Microsoft.

Although open source technologies are increasingly adopted by organizations, there is considerable variation among industries, organizations, and individuals in

terms of their adoption of open source software. Despite their advantages, recent studies also argue that many companies and users are reluctant to adopt open source software in their mainstream activities, primarily because they do not know to what extent they can trust the open source technologies. A survey by Del Bianco, Lavazza, Morasca, and Taibi (2011) has demonstrated that stakeholders at different levels and roles in organizations identify the reliability of open source software as one of top concerns.

The adoption of OSDST also had a massive impact on business practices. Most companies started their data science initiatives with centralized models, creating centers-of-excellence for which they hired data scientists or reassigned relevant employees within the organization. Although this approach enabled dedicated teams with qualified individuals to focus their efforts on data science projects, it causes two problems that hindered taking advantage of the full benefits of data. First, center-of-excellence teams shoulder the burden of nearly all data science projects. Second, depending on one centralized team and ignoring a vast portion of the organization, companies can utilize only a small fraction of the data they possess.

Because most organizations have extensive number of business problems and data-driven decisions to be made at any function, the involvement of knowledge workers need to be considered to exploit data to its fullest extent (Redman & Davenport, 2021). Thus, companies began to consider organizational structures (Figure 1) that allow them to diffuse data science across their organization. Deployed model incorporates small data science teams within business units while diffused model allows large number of individuals to perform analytics work and collaborate with data science experts at the same time. This led to the emergence of the “citizen data scientists” who are described as “individuals who works in fields other than

statistics and analytics yet creates or generates models that incorporate predictive or prescriptive analytics” (Murallie, 2021).

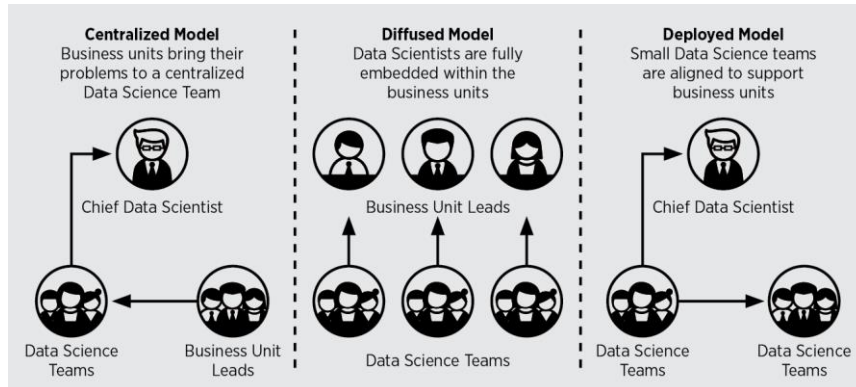


Figure 1. Organizational Data Science Models

Source: Booz Allen Hamilton, 2017

However, this new way of working with data also implies that individuals in business units who generally have non-technical backgrounds also need to adopt and utilize these open source data science technologies. This signifies a substantial change for managers and individuals who are used to work with proprietary technologies for which relying on a major vendor is inherently more reassuring especially when faced with uncertainty and difficulties.

We argue that the risk perceptions about the OSDST has the potential to impend the potential of OSDST adoption by individuals. In fact, a recent study in the limited open source software adoption stream found that perceived risk was a direct indicator of the adoption intention of OSS (Silic & Back, 2015). Therefore, we posit that trust in OSDST is imperative to recognizing the benefits of these technologies as well as utilizing them for business needs.

1.2 Research objectives and questions

To address concerns related to OSDST use, we developed and tested a model that examines the trust-based determinants of OSDST use in organizations. Because of the recency of the research topic, our literature review revealed that there is no empirical study that investigates the trusting beliefs of users neither for OSDST nor for open source software in general. Despite the importance of these technologies for reaping the advantages of data, there is also no empirical research on the factors that would lead to the continued usage of OSDST. Hence, our main research objective is: “to understand the trust-based organizational, social and technological factors that would lead to the continued use of OSDST.”

Specifically, we ask the following research questions:

- What are the roles of different types of trust in the continued usage of the OSDST?
- What is the influence of organizational actors on the general trusting beliefs of users about OSDST and data science community?
- How does the use experience with OSDST (usefulness and satisfaction) influence the utilization of OSDST in current use?
- How does the use experience with OSDST (usefulness and satisfaction) influence the exploration of other OSDST?

1.3 Proposed research model

In this study, we develop a theoretical model that investigates the continued use of OSDST in organizations. We contend that trust in open source data science technologies that are currently in use will have positive influence on the usefulness perceptions and satisfaction with use experiences. These trusting beliefs are hypothesized to be influenced by situational normality and trust in data science

community. We also identify top management championship and co-worker championship of OSDST as the organizational antecedents of both factors.

We point out the critical role of the trust-building mechanisms from several perspectives. We posit that the championship of OSDST by managers and co-workers influence usage experience and behavior through improving the users' positive trusting beliefs on both the data science community and the open source technologies they use at work. To establish these relationships, we rely on prominent theoretical lenses such as IT implementation stage model (Cooper & Zmud, 1990), trust theory (McKnight, Carter, Thatcher, & Clay, 2011), IS continuance model (Bhattacharjee, 2001), and adaptive structuration theory (DeSanctis & Poole, 1994).

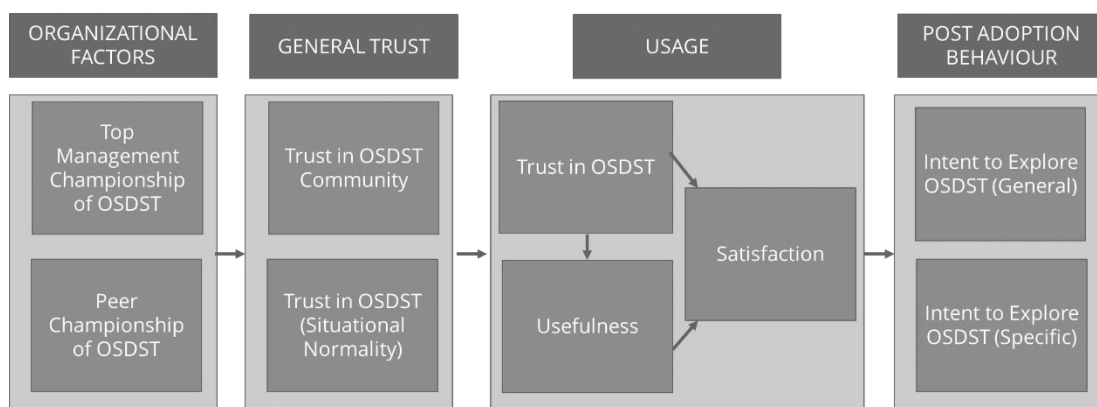


Figure 2. Proposed Research Model

1.4 Contributions

In this study, we have investigated the role of trust-related organizational and technological factors that would influence the postadoption behavior of the users.

First, we developed a model which theorized that having positive trusting beliefs about open source data science technologies is paramount to increasing the usefulness perceptions and developing satisfaction with the use experience of these

products. Then we tested our research model using PLS-SEM with a sample of 173 respondents who performed data analytics across different functions.

To the best of our knowledge, this is the first study that examines the utilization of open source data science and advanced analytics technologies in organizations where most individuals from business units also use open source technologies for the first time. Our findings reveal trust in technology as an immediate determinant of satisfaction. We also found that situational normality and trust in data science community play an important role in shaping users' trusting beliefs in technology.

1.5 Outline of dissertation

This chapter first presents the context around the use of open source data science technologies in organizations, then introduces the research objectives, questions, and model of the study. Chapter 2 provides a review of the relevant literature on technology adoption, continued usage, and trust in the IS literature. We highlight the important findings and identify the gaps in the literature. Chapter 3 develops the research model and hypotheses based on established theories and the findings from the literature review. Chapter 4 describes the methodologies used for the present study. Research design, sampling strategy, research instrument, and data analysis techniques are described in detail. In Chapter 5, results of the data analysis are presented and discussed. Finally, Chapter 6 provides a summary of the findings, highlights the research contributions, acknowledges the limitations of the study, and offers suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

This thesis investigates the continued use of open source data science and advanced analytics technologies in organizations. While open source software can be described as a mature research area, research focus has been mainly on the areas of “motivational factors of individual developers in open source communities” and “the characteristics of specific open source software products and projects” (Fitzgerald, 2006). Because of the public availability of data, “the organizational boundaries between firms and developer communities” also received research attention. However, there is scant research on (1) how data science and advanced analytics technologies are starting to shift and get managed mainly on open source communities, (2) the role of trusting beliefs in usefulness perceptions of technology, and (3) the investigation of continued usage behavior where the users display deep usage and explore the systems actively to maximize the utilization of OSDST. Therefore, we aim to add to the recently growing data science literature (where very few adoption studies exist) and fill in the gaps in the relatively mature open source literature.

Consequently, the literature review will be organized as follows: First, we will review the emergence of data science and advanced analytics and discuss how it has become paramount to organizations. Next, which open source technologies and tools are imperative to different aspects of data science will be discussed. Then, the IT business value stream will be reviewed to corroborate the importance of IT use and satisfaction. Technology adoption literature will be reviewed to identify user behaviors at various stages of IT implementation along with the overarching theories

widely used in literature. Then, we will review the role of trust in technology as well as its influence on IT usage behavior. Finally, the impact of organizational factors on user beliefs and attitudes will be discussed. While doing so, we will borrow from trust theory, adaptive structuration theory (AST), expectation confirmation theory (ECT), and social information processing theory (SIPT) to explain the individual use of IT in organizational settings.

2.1 Data science and advanced analytics

It would not be an overstatement to emphasize the role of Internet for revolutionizing both the access to information and conducting business. First, we witnessed the creation and then the domination of companies such as Google and Yahoo as search engines and infomediaries, while firms such as Amazon and eBay developed novel and successful business models. Although these large Internet players started their operations using vendor-backed legacy technologies, they were also among the first to test their boundaries. The second half of the 2000s witnessed the emergence of social media companies such as Facebook, LinkedIn, and Twitter, with the number of web scale companies rapidly increasing. Many of these large internet players later became the first movers that not only managed to handle these growing data assets but also monetized them using technologies and algorithms that were able to address the data volume challenges at internet scale. This led to the spread of the “big data” term as well as the wide adoption of big data technologies by firms that strived to gain competitive advantage in their markets. (Zillner, Rusitschka, & Skubacz, 2014). Around this time, open source software in general has gained traction as web scale companies kept open sourcing internal projects as part of their sustainability strategies. These companies also successfully used the open source software jointly

developed by the community members, reinforcing the position of open source software as a viable alternative to commercial software. These developments have piqued interest among other traditional enterprises such as banks, telecommunication companies, and retailers to also try and use new generation open source software.

For big data processing, Hadoop, a collection of open source applications that first emerged in 2006 and originally developed for computer clusters, attracted IT managers who needed an efficient method to work with large amounts of data. By January 2008, Hadoop became Apache's top-level project with a diverse and active community as well as large Internet companies such as Last.fm, Facebook, and the New York Times adopting it as active users. In 2011, Yahoo supported the team that developed Hadoop to spin out into a separate Hadoop-focused software company called Hortonworks. By that time, Yahoo's Hadoop infrastructure reached a whopping 42,000 nodes and hundreds of petabytes of storage (Harris, 2013). By 2013, more than half of the Fortune companies were using Hadoop for their big data storage, processing, and analytics purposes (Noyes, 2014). In time, Hadoop became the de facto general-purpose storage and analysis platform for big data. Its dominance has also been recognized by the incumbents of the IT industry such as IBM, Microsoft, EMC and Oracle, all of which either developed their own distributions or provide integration and commercial support for it. The initial success of Hadoop and the related technologies also spurred interest among both more developers and technology investors to move up the technology stack bringing more and more related technologies such as data mining, machine learning and stream processing to the mix. Following Hadoop's success Apache Software Foundation became a magnet for this new wave of projects like Apache Spark, Apache Storm, and Apache Kafka (Woodie, 2016).

The 2010s witnessed the academicians, governments, and research companies starting to explore the big data phenomenon further. In 2011, research company Gartner defined Big Data as “*high-volume, high-velocity and high-variety* information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” (Beyer 2011). In the same year, McKinsey published their now frequently cited report on Big Data, declaring this new phenomenon as the “next frontier for innovation, competition, and productivity” (Manyika et al., 2011). In March 2012, Obama Administration announced the Big Data Research and Development Initiative as part of which Federal Government Institutions committed more than USD \$200 million to big data related research and development. Finally, in October 2012, the big data fire spread to academia in a grand scale: First, Harvard Business Review brought Big Data to its cover dedicating a full issue to the subject. Only a couple of months later, the flagship journal of information systems, MIS Quarterly have published a special issue on Business Intelligence, where several papers have focused on big data. All these developments have further piqued the interest in big data in both business and academia as well as in general public.

In summary, our overview highlights two important trends:

- Big data movement started from the Internet scale companies, later spreading to more mainstream organizations.
 - These Internet-scale firms first developed these technologies to solve their real business needs,
 - Benefited from the use value rather than the sale value,

- Their initial success brought big data to the position of one of the most important potential sources of differentiation for the organizations.
- Open source, particularly the Apache Software Foundation, is identified as not only the current but also the future home of big data software.

While big data refers to information assets themselves, data science refers to “the extraction of actionable knowledge directly from data through a process of discovery or hypothesis formulation and testing” (Edison Data Science Framework, 2017). This definition of data science involves all data activities of accessing, preprocessing, analyzing, and interpreting all sorts of data with the aid of statistical and advanced analytics techniques, tools, and technologies. When it is discovery-based, data science is exploratory in nature and used with the purpose of finding new business opportunities by gaining insights from capturing patterns from data. Such exploratory data analysis is akin to inductive research, first analyzing the data and then understanding the underlying factors before taking the appropriate action. It is mainly utilized by digitally native firms such as Google, Facebook and Netflix, which have huge customer base of millions that generates ever-expanding amounts of big data. On the other hand, data science that is based on hypothesis formulation and testing similar to deductive research, where organizations first identify a business problem, develop initial hypotheses regarding its root causes, and then develop models that generally employ machine learning and advanced analytics techniques.

Apache and Hadoop’s success not only paved way for other open source technologies such as Hbase, CouchDB and Spark on the data processing end, but also

for technologies that allowed users to perform these data science activities. For instance, semi-structured and unstructured data can be stored in and accessed from open source technologies such as PostgreSQL and DuckDB. Extracting external data in websites is possible with BeautifulSoup and Zyte. Data analytics can be performed on development platforms such as Jupyter and Zeppelin while writing Python and R codes for data manipulation and machine learning packages provided by open source communities. KD Nuggets, one of the most popular data science online communities in the world, identify the top data science tools of 2022 (Awan, 2022), confirming the domination open source technologies (Table 1). IT incumbents are playing catch up with these open source developments but managers and developers in software intensive organizations are reluctant to adopt proprietary technologies in this domain due to fear of vendor lock-in and support issues.

Table 1. Top Data Science Tools for 2022

Category	Technology	Source
Database	PostgreSQL	Open Source
	DuckDB	Open Source
Web Scraping	Beautiful Soup	Open Source
	Zyte	Open Source
Data Analytics	Python	Open Source
	R	Open Source
	Julia	Open Source
	Tableau	Proprietary
Machine Learning	FastAI	Open Source
	Scikit-learn	Open Source
	Tensorflow	Open Source
Reporting	Jupyter notebook	Open Source
	Deepnote	Proprietary

When it comes to understanding the adoption of open source data science technologies, Hauge, Ayala, and Conradi's (2010) open source software (OSS) adoption framework provides us a foundation to understand the potentially different ways of deploying OSS. Hauge et al. (2010) points out that while open source software influenced how organizations develop, acquire, use, and commercialize software, adoption of open source software by commercially oriented software-intensive organizations remained relatively less researched area. In an effort to address the challenges of open source software adoption, the authors identify the ways organizations adopt open source software and emphasize that integrating open source components into other software products or systems entails the highest risk and dependence on the open source products (Hauge et al., 2010). Previous research also points out that knowledge barriers such as lack of awareness and difficulty of finding internal and external resources with necessary technical and business know-how seems to play an inhibiting role in open source adoption (Nagy, Yassin, & Bhattacharjee, 2010). Also, a series of interviews as part of a qualitative case study revealed that individuals were hesitant to use new open source technologies because they were not willing to take the risks with them (Dedrick & West, 2003).

In the context of OSDST use, these findings suggests that individuals using OSDST along with other proprietary technology must overcome such risk perceptions as well. Therefore, we believe that the adoption of the open source software, and the role of trust as a risk reducing factor deserve more research attention.

2.2 Continued use and business value of IT

Over the last three decades, research on IS business value have consistently demonstrated that a key ingredient to gaining maximum benefits from IT depends not only on the adoption of technologies but also how individuals utilize them in their workplace practices (Wade & Hulland, 2004). Various IS success models and frameworks suggest that effective use of IT is a precursor of individual performance, which in turn is one of the determinants of organizational performance. For instance, Soh and Markus (1995) present a process theory that examines the linkage between IT investments and business, where they identify the appropriate use of technology as the main determinant of gaining positive impacts of the investments as well as the overall organizational performance.

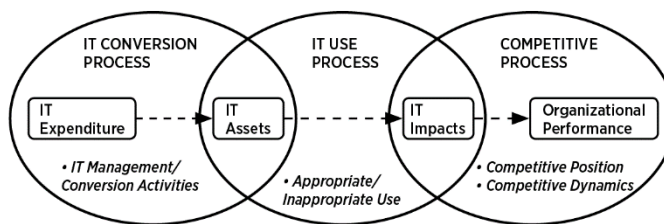


Figure 3. IS Process Model

Source: Soh & Markus, 1995

In their seminal IS Success Model studies, DeLone and McLean (1992, 2003) explain the relationships between the most critical dimensions of IS success in organizations, where “IT Adoption/Use” and “User Satisfaction” are identified as the two immediate determinants of “Net Benefits” which reflects the contribution of the system to the performance of the individuals and organizations. The authors emphasize IT use not in terms of usage quantity but the depth of the utilization of its capabilities. Similarly, Schryen’s (2013) IT Business Value model suggests that

individual IT usage would improve an organization's business performance through its direct effect on business processes.

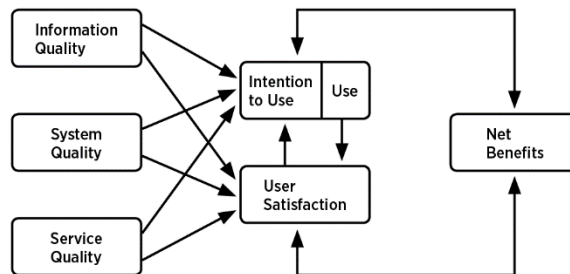


Figure 4. Updated IS Success Model

Source: DeLone & McLean, 2003

Overall these studies together concur that:

- The benefits of technologies can only be achieved if IT is fully utilized and integrated into one's work in a meaningful manner.
- Understanding the processes and factors that influence continuation and utilization of IT is crucial for both research and practice.

In the context of this study, the utilization of OSDST is viewed critical for organizations to reap the potential benefits from data. This is done by extracting insights and taking actions in order to gain competitive advantage over rival firms. Consider the most widely used open source programming languages such as Python and R. Employees may initially adopt them, learn the fundamentals of coding, and understand the array of packages available. Unless these users continue to utilize these technologies and make them a routinized part of their work lives or put effort to explore their capabilities, the initial acceptance is likely to fade away. As a result, either users will quit using them or at best case will not utilize their capabilities fully, impeding successful data-driven decision making efforts eventually. Therefore, we

turn to the IT adoption literature to review the theoretical and empirical research that sheds light on continued use of technology.

2.3 IT adoption and implementation in organizations

We base our understanding of IT adoption and use in organizations on the seminal implementation stage model of Cooper & Zmud (1990). While the stages may overlap at times, the IT implementation process involves six subsequent stages.

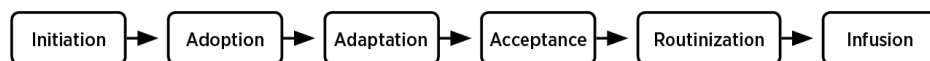


Figure 5. IT Implementation Model

Source: Cooper & Zmud, 1990

First three stages (initiation, adoption, adaptation) represent the primary adoption of technologies by organizational decision makers. The latter three stages (acceptance, routinization, infusion) refer to the secondary adoption and use of technologies by individuals in organization, although adaptation stage plays an important role for technology users since this is when they get trained for the new technology. Therefore, IS research considers organizational level adoption as the process that begins when the organization gathers information about the technology under consideration and concludes when it acquires and the technology. On the other hand, individual-level adoption examines the process that begins with the introduction of the new technology and its continued use (Jeyaraj, Rottman, & Lacity, 2006).

The IT implementation model consists of:

- Initiation: The stage in which
 - the organization faces a business problem or opportunity,
 - gains information about the potential IT solutions, and
 - eventually finds a match between the technology and its application.
- Adoption: The stage in which
 - the organization reaches an agreement for the IT solution, and
 - invests organizational resources for its implementation.
- Adaptation: The stage in which
 - the organization develops or purchases new IT,
 - develops and revises organizational procedures, and
 - provides necessary trainings for employees.
- Acceptance: The stage in which
 - individuals are encouraged to use the new technology, and
 - the new technology is used for performing organizational work.
- Routinization: The stage in which
 - usage is encouraged as a normal activity, and
 - new IT is perceived as something normal at workplace.
- Infusion: The stage in which
 - IT is used in a more inclusive and cohesive manner to reach its full potential.

Table 2. Summary of IT Implementation Stages

Stage	Name	Implementation Level	Definition
1	Initiation	Organization	Finding new IT that best fits the business needs
2	Adoption	Organization	Organizational resources are committed for new IT
3	Adaptation	Organization & Individual	Organization rollouts new IT. Training provided for users.
4	Acceptance	Individual (Initial Adoption)	Users' initial commitment to use the system
5	Routinization	Individual (Post-Adoption)	Use of IT becomes part of the user's behavioral routine
6	Infusion	Individual (Post-Adoption)	Use of IT is fully utilized at work.

2.4 Organizational level adoption

An overview of the rich organizational level adoption stream can be found in excellent literature reviews and meta-analyses by Oliveira and Martins (2011), Hameed, Counsell, and Swift (2012), and Lauterbach and Mueller (2014). The findings of these studies reveal that organizational level adoption studies follow two different approaches to examining IT adoption: One stream focuses on the process of how technological innovations permeate organizations, explaining the process that begin with the adoption and followed by its implementation. This line of research employs mostly longitudinal qualitative case research, hence providing insight for at least some phases of technology implementation. The second stream mainly uses cross-sectional survey methodologies and focuses on understanding the organizational and technological factors that determine the innovation adoption.

In terms of theoretical perspectives, Diffusion of Innovations (Rogers, 1995) and Technology Organization Environment Framework (Tornatzky & Fleischer, 1990) are two of the most widely used theories of organizational adoption that

complement each other. Diffusion of Innovations (DOI) is essentially a process theory that provides an explanation for the factors that determine at which rate technologies spread over time. In the context of organizational-level technology adoption, Rogers (1995) identifies the innovation characteristics such as relative advantage, complexity, compatibility, observability, and trialability.

While Rogers' model is at a conceptual level, empirical studies that investigate the influence of these factors derive their measures from Moore and Benbasat's (1991) study, which developed and validated scales specifically for the IT context. On the other hand, Technology Organization Environment Framework (TOE) explains how the firm context, with technological, organizational, and environmental set of factors, influence the adoption and implementation of technologies (Tornatzky & Fleischer, 1990). While DOI provides a strong explanation for the technological (T) and organizational (O) determinants of adoption, TOE has the ability to explain IT adoption better by adding environmental factors (E) as well.

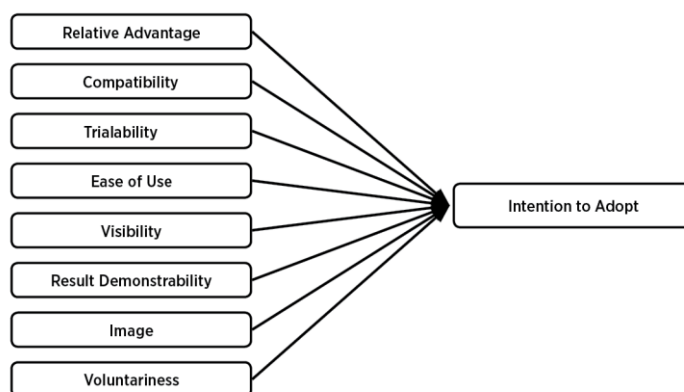


Figure 6. DOI Model for IT Adoption

Source: Moore & Benbasat, 1991

Studies that adopt TOE identify the relevant direct and moderating factors depending on the context while developing their models. For instance, Zhu, Kraemer, and Dedrick (2004) demonstrated that the influence of TOE factors depended on the country characteristics. They found that the regulatory environment played a more significant role in underdeveloped countries while the firm size was a deciding factor in developed countries in adoption decisions. Borgman, Bahli, Heier, and Schewski (2013) found that governance structures and processes of firms moderate the relationship between TOE factors and adoption decisions of cloud computing technologies. There are also studies that combine DOI and TOE in an attempt to improve the explanation of the variation in adoption. For instance, Oliveira et al. (2014) examine the role of both TOE factors and innovation characteristics together in the context of cloud adoption. In addition to IT adoption, TOE framework has also been applied to studies that examine IT value of technologies (e.g., Kuan & Chau, 2001).

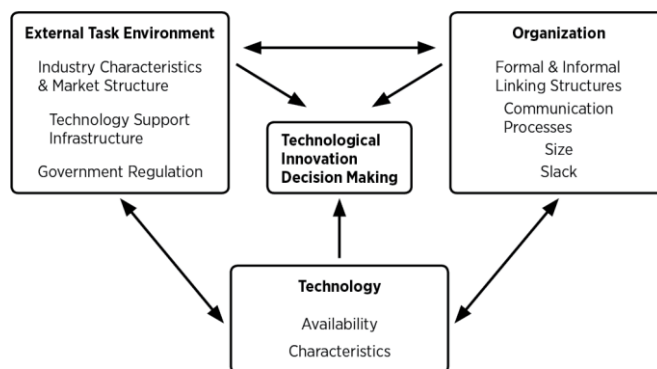


Figure 7. TOE Framework

Source: Tornatzky & Fleischer, 1990

2.5 Individual level adoption

Individual level adoption studies measure the initial user acceptance and the actual use of technologies (Hameed et al., 2012). In terms of Cooper and Zmud's (1990) IS Implementation Model, these studies fall into two major streams: First stream is comprised of studies that examine the processes and factors that lead to the initial adoption of IT at the "Acceptance" stage. Second stream consists of research that investigate the post-adoption behaviors of users and study the factors that cause a user to continue or discontinue a technology as well as utilize it extensively to improve individual and organizational performance. These studies fall into "Routinization" and "Infusion" phases, although the line between these stages is blurrier than the others.

2.5.1 Initial acceptance

The initial acceptance of IT is one of the most studied areas in IS research. In general, these studies build their models on the established theories found in social psychology, a discipline that hosts some of the most influential theories that explain human behavior. One of these theories, Theory of Reasoned Action (TRA; Fishbein & Ajzen, 1975), posits that an individual's salient beliefs and evaluations about a behavioral outcome would influence their attitudes towards that behavior. These attitudes then influence their intentions, which in turn leads to actual behavior. The defining research for IT acceptance is the set of studies conducted by Fred Davis and his colleagues (Davis, 1986; Davis, 1989; Davis, Davis, Bagozzi, & Warshaw, 1989).

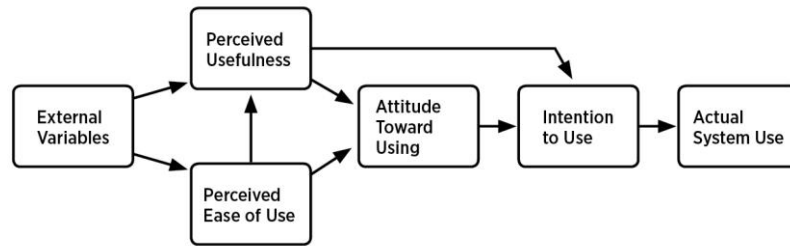


Figure 8. Technology Acceptance Model

Source: Davis et al., 1989

Davis's Technology Acceptance Model adopted the Theory of Reasoned Action and tailored it for the IT context. The model identifies the two salient beliefs about the technology, perceived usefulness (the degree to which an individual believes that using the technology would enhance performance at work) and perceived ease of use (the degree to which an individual believes that using the technology would be free of effort) (Davis, 1986; Davis, 1989). In a study in which he examined the acceptance of a word processor, Davis (1989) compared his model with the original TRA, and found that TAM has a significantly better explanatory power in explaining technology acceptance. Following years have witnessed numerous research studies that used TAM as a baseline, determining the external factors that influence the salient beliefs of usefulness and ease of use. While other theories of IT acceptance have also been developed and tested (see Table 4), the next breakthrough acceptance model has emerged in 2003 with the Unified Theory of Acceptance and Use of Technology (Venkatesh, Morris, Davis, & Davis, 2003)

Table 3. Selected Technology Acceptance Models

Theory	Origin	Contribution
Technology Acceptance Model	Davis, 1986; Davis, 1989	Extends TRA by identifying perceived usefulness and perceived ease of use as salient beliefs that shape attitudes and use intention
Technology Acceptance Model 2	Venkatesh and Davis, 2000	Extends TAM by including social influences that predict perceived usefulness
Motivational Model	Davis, Bagozzi, and Warshaw, 1992	Examines the role of extrinsic and intrinsic motivational factors on acceptance
Decomposed Theory of Planned Behavior	Taylor and Todd, 1995	Introduces perceived behavioral control as constraints on use behavior
Unified Theory of Acceptance and Use of Technology	Venkatesh et al., 2003	Develops and tests a combinative model of previous acceptance theories

Unified Theory of Acceptance and Use of Technology (UTAUT) was developed with the motivation of creating a model that could explain the use behavior for more complex and sophisticated technologies. The authors compared and tested eight widely used acceptance models in a longitudinal study held in four organizations from four industries, finding strong support for the new model. UTAUT has refined the usefulness and ease of use definitions and scales, introduced facilitating conditions and social influence and as direct antecedents of behavioral intention as well as age, gender, and experience as moderating factors (Venkatesh et al., 2003). Overall, the UTAUT model explained an impressive 70% of the variation in use intention, and has contested TAM as researchers' preferred IT acceptance model.

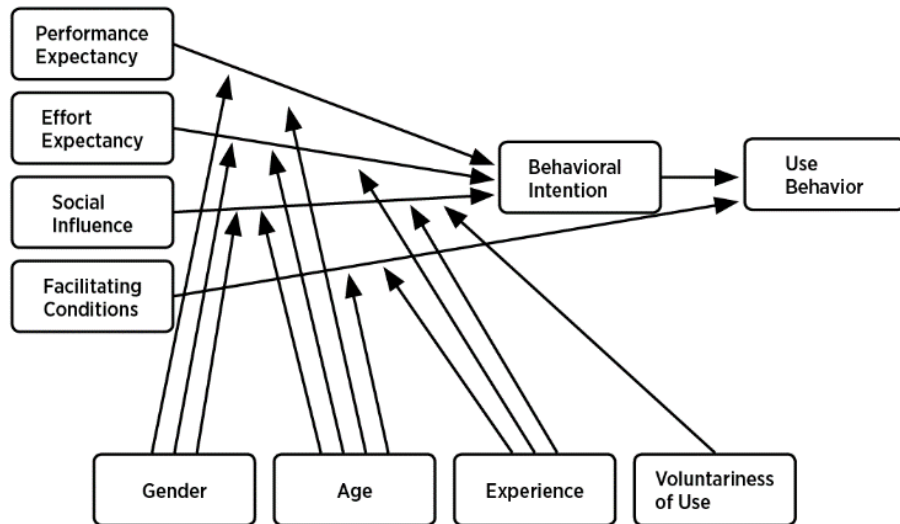


Figure 9. UTAUT

Source: Venkatesh et al., 2003

The TAM and UTAUT models of technology acceptance have had massive impact on not only the IS field, but many disciplines that examined individual-level IT adoption. A Web of Science search demonstrates this impact: As of 2022, TAM and UTAUT have been cited nearly by 8,600 and 2,100 times. Table 4 lists some of the most cited studies that adapted these models for different contexts.

Table 4. TAM and UTAUT in Various Contexts

Theory	Context	Discipline	Reference
TAM	Online shopping	Information Systems	Gefen, Karahanna, and Straub, 2003
TAM	Health IT	Health	Holden and Karsh, 2010
TAM	E-learning	Education	Park, 2009
TAM	Online banking	Information Systems	Pikkarainen, Pikkarainen, Karjaluoto, and Pahnla, 2004
TAM	Handheld internet devices	Business	Bruner and Kumar, 2005
UTAUT	Mobile banking	Information Systems	Zhou and Wang, 2010
UTAUT	Virtual learning environment	Education	Van Raaij and Schepers, 2008
UTAUT	Health IT	Health	Kijsanayotin, Pannarunothai, and Speedie, 2009
UTAUT	Online shopping	Tourism	San Martin and Herrero, 2012
UTAUT	Blockchain	Operations Management	Francisco and Swanson, 2018

2.5.2 Continued use

Because information technologies in organizations are mostly underutilized, investigating only the initial acceptance may not be adequate to understand the whole usage process of these systems. Continued use points to the “routinization” and “infusion” stages of the IT implementation model. Routinization is important because it indicates that the user perceives the new technology as a part of business as usual, i.e., normalizes the use of technology at workplace practices. If a user has reached the infusion stage, it means that he or she utilizes more features, develop links between different work tasks, and even performing tasks that was never acknowledged before. Therefore, this stage is where the real value of usage lies since the technology is utilized to its maximum value, the aim for technologies that are adopted by the organizations (Agarwal, Sambamurthy, Stair, 2000; Cooper & Zmud, 1990; Jasperson, Carter, & Zmud, 2005). Research that studies IT continued use and examine these two stages represent the “post-adoption” research stream.

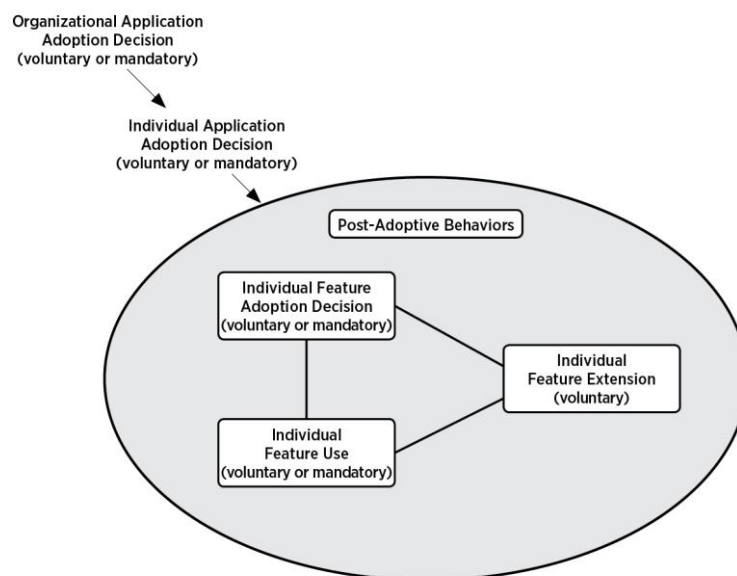


Figure 10. Feature-Centric View of IT Adoption and Use

Source: Jasperson et al., 2005

Although TAM and UTAUT are developed specifically for initial adoption, they are not suited well to predict the continuance behavior since they do not explain how beliefs and intentions change over time (Kim & Malhotra, 2005). IS continuance model (Bhattacharjee, 2001) and the feature centric view of IT adoption use (Jasperson et al., 2005) are two theoretical frameworks that address these problems. We adopt these perspectives to explain our post adoptive behavior of interest in the present study.

To explain the importance of technology exploration, we rely on the “feature-centric view of IT adoption and use” (Jasperson et al., 2005). Post-adoptive behavior is defined as a set of users’ feature adoption, feature use, and feature extension behaviors after the new IT is made available to them and applied in undertaking their work tasks (Jasperson et al.,2005), The theory emphasizes the feature extension behavior, which is associated with exploring the technology further and is an indicator of reaching the infusion stage (Jasperson et al.,2005). In line with the feature-centric view, intent to explore IT has been studied in several research that examined its determinants (Lippert & Forman, 2005; Nambisan, Agarwal, & Tanniru, 1999; Thatcher, McKnight, Baker, Arsal, & Roberts, 2011). Intent to explore a technology comprises of two components: (a) willingness to use a variety of features of a technology and (b) to actively think about how to incorporate that technology into work (Nambisan et al., 1999, Thatcher et al., 2011).

The feature-centric view of IT adoption also identifies a user’s satisfaction with the initially adopted system as the “sense-making of the technology” and one of the most critical factors in the post-adoption usage stage (Jasperson et al., 2005). When studying user satisfaction, IS continuance model provides us the theoretical

framework that justifies our choice of OSDST satisfaction as the immediate determinant of exploration behavior.

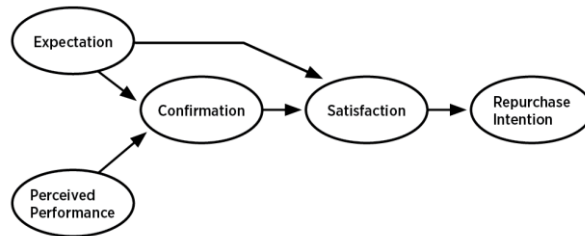


Figure 11. Expectation-Confirmation Theory

Source: Bhattacharjee, 2001

The premise of user satisfaction can be found in the expectation-confirmation theory (Oliver, 1980). Expectation-Confirmation Theory (ECT) was originally developed in the marketing discipline to explain the repurchase intentions of the customers. It posits that the customers who are satisfied with the products or the vendors will continue to buy the same products or shop from the same vendor. Satisfaction is dependent on the perceived performance (like perceived usefulness in TAM and performance expectancy in UTAUT) and prior expectations.

The IS continuance model (Bhattacharjee, 2001) builds on the foundations of ECT, and theorizes that users first form an initial pre-usage expectation about the technology and then start experiencing the usage over time, from which they develop post-usage perceptions about the IT. The difference between the observed performance and the pre-usage beliefs form disconfirmation. If the gap between performance and pre-usage beliefs are positive, the user is expected to be satisfied with it, and continue using the software. On the other hand, if the gap is negative, then the user is expected to be dissatisfied with it, leading to the discontinuance of usage.

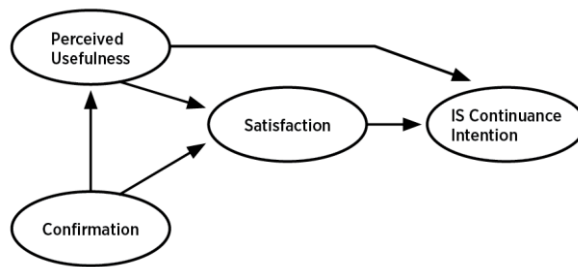


Figure 12. IS Continuance Model

Source: Bhattacharjee, 2001

The model is applied extensively in postadoption research. These studies reveal that satisfaction has consistently been found to be associated with post-adoption behavior. For example, Thong, Hong, and Tam's (2006) study revealed satisfaction as the strongest indicator of continuance intention among the other proposed determinants i.e., perceived usefulness, ease of use, and enjoyment. In a study where habit was suggested as a determinant of IS continuance, satisfaction was found to influence both habitual use as well as IS continuance intention (Limayem, Hirt, & Cheung, 2007). In the context of e-government technologies, Venkatesh, Thong, Chan, Hu, and Brown (2011) found that satisfaction positively influenced a user's post-use attitudes towards continuance behavior. While these studies measured the intent to continue IT, others have adopted a feature-centric view and selected the dependent variable as "extended use". They defined extended use as "individuals using more of the technology's features in order to accommodate a more comprehensive set of work tasks" (Saga & Zmud, 1994). For instance, Hsieh and Wang (2007) tested the IS continuance model in the context of ERP use in China and found satisfaction as a determinant of extended use intention. Consistent with the IS continuance model and Hsieh and Wang's findings, we argue that satisfaction will

influence a user's intent to explore OSDST, a construct that is similar to extended use. Next, we turn to trust theory as we posit that trust-related factors (trust in OSDST, trust in OSDST community, and situation normality) are central to understanding the usefulness perceptions and satisfaction.

2.6 Trust in technology

Trust is central in situations where one feels the presence of risk, uncertainty, and interdependence (Mayer, Davis, & Schoorman, 1995). In recent years, trust in technology has emerged as a critical factor in the IS research streams that study technology adoption and e-commerce. In both streams, both the determinants and the outcomes of trusting beliefs are examined in different ways, varying on the technology and the context. As in interpersonal trust or trust in vendors, developing positive trusting beliefs about technology is vital because it allows people to reduce their risk beliefs associated with technology use (McKnight, Choudhury, & Kacmar, 2002).

Research within the e-commerce domain examine user trusting beliefs in technology on the initial adoption or use of various online artifacts such as websites and recommendations agents since individuals perceive them as valid social actors like other interpersonal relationships (Moore & Grover, 2010; Wang & Benbasat, 2005; Zimmer, Arsal, Al-Marzouq, Moore, & Grover, 2010). In return, positive trusting beliefs have been found to influence consumer beliefs, attitudes, and behavior in online settings (Gefen et al., 2003; Lee, Kang, & McKnight, 2007; Vance, Elie-Dit-Cosaque, & Straub, 2008). Trust is discovered to be so important that its influence on shopping intentions were found to be more significant than cheaper prices (Jarvenpaa, Tractinsky, & Vitale, 2000).

Research that examines the role of trust in technology in use of IT artifacts demonstrate similar findings both for initial and post adoption. For example, Thatcher et al. (2011) demonstrated that user trusting beliefs in knowledge management systems was positively related to exploration intentions. In a study that examined the utilization of Microsoft Excel and Access applications, McKnight et al. (2011) found that users who had favorable trusting beliefs were expected to explore and utilize various features of these technologies. In a more recent study, Tams et al. (2008) theorized that the influence of trust in technology on post adoption behavior was positive again, adding the role of self-efficacy beliefs as a mediating factor.

Trusting beliefs in OSDST is particularly important because the reliability of the final products usually depends on the underlying open source components and involves a great deal of risk (Auger, 2016). In the present study, we position trust in OSDST as an immediate antecedent of perceived usefulness and satisfaction. Relying on the trust theory, we also identify situational normality and trust in OSDST community as the determinants of trust in OSDST.

We base our trust conceptualizations on McKnight, Carter, and Clay (2009) and McKnight et al.'s (2011) trust models that adapt the generic trust types to trust in IT and develops formal definitions of the construct. Trust in IT means “relying or depending on specific information systems such as software or applications” (McKnight et al., 2005). While previous research demonstrated that people in fact trust in technologies, it is inherently different from trust in people since technologies do not have moral or volitional attributes (McKnight et al., 2009). Therefore, trusting beliefs should reflect the capabilities of the technologies.

In a series of studies that examined trust in technology, McKnight et al. (2009, 2011) first conceptualized the user beliefs in technological capabilities and

then developed and validated the construct measures. According to trust theory, an individual's trusting beliefs are reflected in a combination of one's trusting belief in the competence, benevolence, and integrity of the trustees in the interpersonal situations. In technology context, trust in functionality beliefs takes the place of competence, helpfulness replaces benevolence, and predictability substitutes integrity. The definitions of the constructs and the reasoning behind their differences are summarized in Table 5. For this study, we adapt this conceptualization. Accordingly, we treat trust as a second-order construct that consists of three reflective first-order constructs that measure trusting beliefs in the functionality, helpfulness, and predictability of the open source data science and advanced analytics technologies that the users are experiencing.

Table 5. Trusting Beliefs: Dimensions and Conceptualizations

Interpersonal Trust	Trust in Technology	Differences
Competence: The belief in the ability of the trustee to do what the one needs	Functionality: The belief that the system has the capability, functions, or features to do for one what one needs to be done	People have skills Systems have features
Benevolence: The belief that the trustee cares for one and has one's interests at heart	Helpfulness: The belief a system will provide adequate and responsive aid	People have altruistic motives Systems have help functions
Integrity: The belief that the trustee acts according to accepted principles such as keeping promises	Predictability: The belief that the system acts consistently, and its behavior can be forecast	People have principles and keep their promises Systems act predictable or consistent in how it works
Source: Mayer et al. (1995)	Source: McKnight et al. (2002) McKnight et al. (2011) Thatcher et al. (2011)	Source: Thatcher et al. (2011)

Next, we turn to the nomological network of trust in technology (Figure 13) that helps our understanding of how institutional trusting beliefs shape a user's trust

in OSDST they are using. Our two constructs, situational normality, and trust in OSDST communities, represent these institutional beliefs, and represent the “general trust” block of our research model.

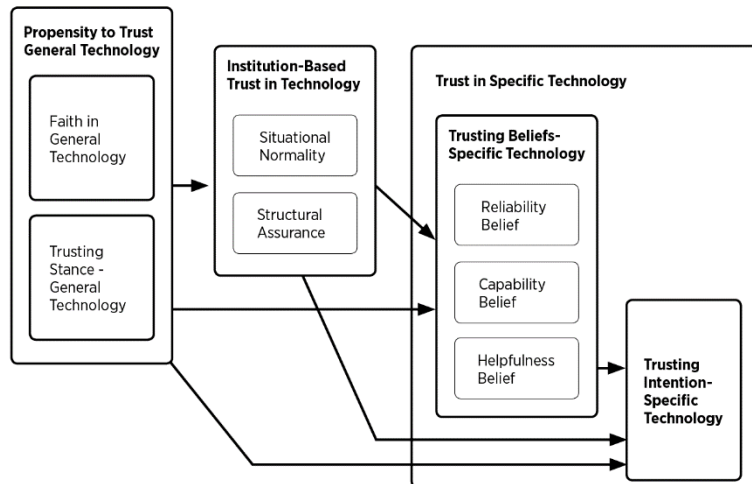


Figure 13. Nomological Network of Trust in Technology

Source: McKnight et al., 2009

Trust theory views situational normality beliefs as perception of the situation as normal or favorable. In return these beliefs can be extended to something new in that situation (McKnight, Cummings, & Chervany, 1998). Situational normality in technology context refers to “the belief that using a specific class of technologies in a new way is normal and comfortable within a specific setting” (McKnight & Chervany, 2001). Situational normality has been proposed and found to play a significant role in developing trusting beliefs when something new occurs or comes up in the situation where individuals have already developed such favorable beliefs. For instance, in their seminal paper that developed a conceptual trust model that examined the formation of trust in organizations, situational normality has been theorized as an immediate determinant through helping employees becoming

comfortable in their roles in the organization (McKnight et al., 1998). Empirical research supports the theory: Gefen (2003), in his seminal paper on understanding online shopping intentions found that situation normality (a customer's expectation that the web site looks normal and similar to others online) had the strongest influence on shaping a customer's trust in a specific online seller. McKnight et al. (2011) demonstrated that situational normality beliefs in spreadsheet applications had a positive association with trust in a specific spreadsheet application, Microsoft Excel.

In the context of OSDST use, we position situation normality consistent with the trust theory. Situational normality is an important determinant of trust in OSDST beliefs because if the users do not buy into the idea that it is the new preferred way of performing data-related tasks at work, their risk perceptions of using a technology will overshadow its benefits. The users are going to be less likely to adopt or utilize open source technologies because it will not be worth taking risks when the organization does not create a supportive environment that convinces and encourages the users that open source data science technologies is the new way for data science.

In this study, we position "trust in OSDST community" as a proxy for structural assurances in the trust in technology model. Structural assurance refers to the supporting infrastructure such IT support and contractual guarantees that provides the adequate support to ensure the utilization of technologies (McKnight et al., 2011). However, open source software developments do not provide the support and warranties like proprietary technologies. However, although they do not provide commercial support, they close this gap with the support of the community formed by the developers and people who use these technologies (Malone & Wolski, 2020). Some of these communities are Python/R forums, Kaggle, Stack Overflow, and Data

Science Central. Therefore, developing and maintaining structural assurances in the context of OSDST adoption depends on the presence, quality, and the helpfulness of the community. In this study, we conceptualize trust in OSDST as a user's willingness to depend on and having favorable beliefs about the attributes of the OSDST community.

Next, we turn to technology championship in organizations since we propose top management and co-worker championship as the determinants of situational normality and trust in OSDST community.

2.7 IT championship in organizations

The influence of top management and co-worker championship in IT adoption is best explained through the lens of adaptive structuration theory (AST), which is based on Anthony Giddens' structuration theory used to explain “the production and reproduction of the social systems through members’ use of rules and resources in interaction” (Giddens, 1994). AST is developed by DeSanctis and Poole (1994) to explain the complex and social nature of the IT adoption and use. AST argues that IT utilization cannot be predicted by focusing on the technological characteristics alone and employee behavior is a result of the intertwining factors such as organizational and social structures, rules, procedures, and resources provided by technologies (DeSanctis & Poole, 1994). The theory posits that neither these factors nor their interactions are static. In other words, all institutional, technological, group, and individual entities continuously influence and shape each other over time (Poole & DeSanctis, 2004).

Adaptive structuration theory argues that new technologies would be more successfully implemented when they are actively promoted and endorsed by

influential individuals within the organization (Orlikowski, Yates, Okamura, & Fujimoto, 1995; Orlikowski, 2000). Therefore, IS research that adapted AST identifies top management championship as one of the critical determinants of technology adoption in organizations. Championship activities are important because the beliefs and actions of managers define the organizational norms and values. Through their beliefs in the business benefits of technologies, top management can legitimize their adoption and use. When belief signals are coupled with their active participation, they also offer visions, strategies, and guidelines to the rest of the organization. The stronger these beliefs and participation activities, the stronger signals will be received by managers, business units, and individual employees (Chatterjee, Grewal, & Sambamurthy, 2002).

Adaptive structuration theory also suits well for the context of OSDST use in organizations because the nature of data science and advanced analytics necessitate the involvement of individuals in teamwork rather than working alone. Thus, user beliefs and perceptions about OSDST are likely to be shaped by a combination of management, peers, technologies, and resources that facilitate their utilization. Through their beliefs, top management can emphasize the necessity of utilizing OSDST in business activities to improve organizational performance and gain competitive advantage over their rivals. Moreover, through their participation, they can provide strategies, goals and standards which act as guidelines. Legitimizing the use of these technologies with such championship activities is likely to cause OSDST users to mitigate their risk perceptions of working with open source technologies, develop more favorable attitudes towards using them, and feel that using a particular class of technologies is the normal way of performing data-related work.

2.8 Individual adoption of open source software in IS research

Although most of the IS research on open source software focus on their adoption at the organizational level and the motivation of developers to contribute to open source, there are several studies that examine OSS adoption at the individual level.

A few of these studies in this research stream adopted the TAM constructs. Gallego, Luna, and Bueno's user acceptance model of OSS (2008) is the first study to examine the factors that influenced OSS use. The authors confirmed that the salient user beliefs in TAM, perceived usefulness and perceived ease of use, also had a significant association with intention to use Linux operation system. They also identified OSS system quality and OSS flexibility as the most important determinants of perceived usefulness perceived ease of use respectively. In an extension of this study, Gallego, Bueno, Racero, and Noyes (2015) emphasized the impact of training. This time, they found that user training had a positive influence on perceived usefulness while the support of the trainers had a direct impact on use behaviour.

On the other hand, Li, Tan, Xu, and Teo (2011) used motivation and self-determination theories to develop two separate models to understand if there are different factors for the adoption and non-adoption of open source software. Their study demonstrated that the expected value of OSS played the most significant role in the decision of adoption. Individuals who believed in the benefits of OSS were found to be intrinsically and extrinsically motivated. On the other hand, individuals who believed that using OSS would not lead to desired outcomes shied away from adopting them as a result.

In a qualitative case study that was theoretically based on Innovation Diffusion Theory (Rogers, 1995) and conducted in two public administrative organizations, Rossi, Russo, and Succi (2012) acknowledged the role of the

management-user alignment. They found that the gap about the perceived benefits of open source between the management and individuals inhibited the migration of OSS. The study points out to the importance of the alignment between managers and users as well as the development of a training strategy for the users.

Though our literature review identified a few studies that examined open source software adoption at the individual level, we found no study that investigated open source adoption in the context of data science and advanced analytics.

Table 6. Construct Definitions

Construct	Definition
Top Management Championship of OSDST (Belief)	User perception of senior management's beliefs in the operational and strategic benefits of OSDST for the organization.
Top Management Championship of OSDST (Participation)	User perception of senior management's active participation in the adoption of OSDST
Co worker Championship of OSDST (Belief)	User perception of co-workers' beliefs in the operational and strategic benefits of OSDST for the department.
Co worker Championship of OSDST (Participation)	User perception of co-workers' active participation in the adoption of OSDST
Trust in OSDST Community	The extent a user is willing to depend on and has favorable beliefs about the attributes of the OSDST community
OSDST Situational Normality	The belief that success with OSDST is likely because an individual feels normal and comfortable using them at the workplace.
Trust in OSDST (Functionality)	The belief that the OSDST has the capability, functions, or features for the users to perform the tasks that needs to be done.
Trust in OSDST (Helpfulness)	The belief that the OSDST provides adequate and responsive aid.
Trust in OSDST (Predictability)	The belief that the OSDST acts consistently, and its behavior can be forecast.
Perceived Usefulness	The degree to which a user believes that using OSDST enhances his or her job performance
Satisfaction	User's feelings about the OSDST that he/she has experienced.
Intent to Explore OSDST(Current)	User purpose and motivation to innovate with the current OSDST in use.
Intent to Explore OSDST (Others)	User motivation to investigate and experiment with OSDST that they are currently not adopted.

CHAPTER 3

HYPOTHESES DEVELOPMENT

Following our literature review on IT adoption, trust, and technology championship in organizations, we develop our hypotheses. The research model is shown in Figure 14.

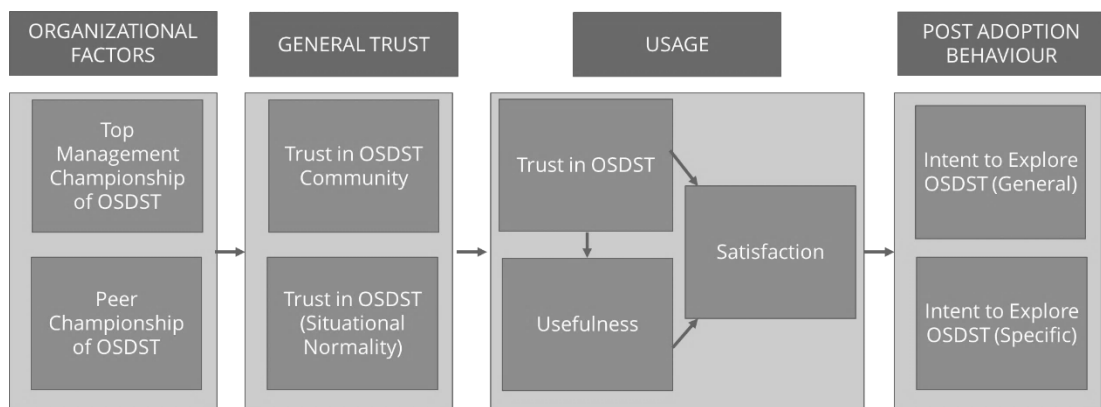


Figure 14. Research Model and Hypotheses

We propose that top management championship and co-worker championship of OSDST positively influence a user's perceptions of trust in OSDST community and OSDST situational normality. In a study that investigated the organizational assimilation of web technologies, Chatterjee et al. (2002) found out that top management championship had a significant and positive impact on the assimilation process. Sharma, Singh, Jones, Kraus, and Dwivedi (2022) investigated the adoption of agile management practices in SMEs and found that top management championship was positively related to the adoption intention. In a study conducted in health industry, Gopalakrishna-Remani, Jones, and Camp (2019) found that top management championship positively influenced the adoption of electronic medical reports in hospitals.

Social information processing theory (Salancik & Pfeffer, 1978) also offers additional explanation for the influence of co-worker championship of OSDST. The theory posits that the workplace is a social context where the beliefs, attitudes, and actions of individuals are influenced by the information conveyed by others at the workplace can also influence their cognition about technologies.

Therefore, we expect strong championship of OSDST from both management and co-workers to help users develop beliefs that using open source technologies is normal and even favorable. Because there is no traditional technical support as provided by proprietary technologies, championship efforts should also influence user beliefs that data science community would provide the necessary structural assurances.

H1. Top management championship of OSDST will have a positive influence on OSDST situational normality.

H2. Top management championship of OSDST will have a positive influence on trust in OSDST community.

H3. Co-worker championship of OSDST will have a positive influence on OSDST situational normality.

H4. Co-worker championship of OSDST will have a positive influence on trust in OSDST Community.

Trust in OSDST community is expected to influence user perceptions of situational normality and trust in OSDST. It acts as a structural assurance for OSDST users in the shape of providing resources and help when the users seek information or assistance. When sufficient support exists, it can ensure the effective use of a technology (McKnight et al., 2011). IS literature that examines the influence of IT

support on technology usage point out to the IT staff characteristics that would lead to the adoption or continued usage of technologies. For instance, Nelson (1991) demonstrates the importance of not only the technical skills of an IT support but also their ability to deal with people who use that technology. In a later study, Thatcher et al. (2011) demonstrated that the influence of a user's trust in IT staff had a positive association with trust in technology.

In the present study, we argue that OSDST community plays a similar role to that of an IT personnel or vendor support staff that provides assurances for the users. Successful open source communities do not only develop the technology and make it available for the community. Their members also act as support staff—demonstrating altruistic behavior by helping other members when dealing with implementation problems or when individuals are using them. When users believe in the reliability and helpfulness of the IT staff, they develop positive attitudes and behaviors towards the technologies that they use for their work tasks (Nelson, 1991). In a similar way, OSDST users who trust the OSDST community should be able perceive positive beliefs about the software developed in that community. Hence, we expect user trust in the OSDST community to be positively related to user beliefs of situational normality and trust in OSDST.

H5: Trust in OSDST Community will have a positive influence on situational normality.

H6: Trust in OSDST Community will have a positive influence on the trust in OSDST.

Empirical research provides evidence to the relationship between situational normality and trust in technology in different contexts. In earlier stages of e-commerce research, Gefen et al. (2003) found that situational normality beliefs about websites would lead to high trusting beliefs about an online vendor. Pennington, Wilcox, and Grover (2003) emphasized the importance of situational normality in reducing the uncertainty that increase the risk perceptions, which would undermine transactions with specific vendors. In two separate studies, Wingreen and Baglione (2005) and Wingreen, Mazey, Baglione, and Storholm (2019) also found that situational normality beliefs about online shopping improved trusting beliefs about specific websites and vendors.

Similar results are reported when the focal point of trust is an IT artifact. McKnight et al. (2011) examined the role of situation normality for postadoption exploration behavior for Excel users and found that user beliefs about the Excel application were significantly improved when users have favorable beliefs about working with spreadsheet products in general.

Following the trust theory and the empirical evidence from prior literature, we posit that if an individual believes that the use of open source data science technologies has become a normal and favorable way of performing data-related tasks, these beliefs will extend to trusting beliefs with a specific open source technology within the same class. Therefore, we hypothesize that:

H7. Situational normality will have a positive influence on trust in OSDST.

According to Technology Acceptance Model (Davis 1986, 1989), perceived usefulness is one of the two salient beliefs about technology that would have the

most significant influence on adoption and usage intentions. Trust theory suggest that when users have positive beliefs about the functionality, helpfulness, and predictability of the technologies they are utilizing, they will also develop positive beliefs about their usefulness perceptions. Previous literature demonstrates that trust in technology can induce technology related salient beliefs in different contexts such as application software and e-commerce websites (e.g., Gefen et al., 2003; Pavlou, 2003; Thatcher et al., 2011). Because relying on OSDST software (as well as other complex IT) involves uncertainty and risk, developing trusting beliefs in them is critical. For instance, Benbasat and Wang (2005) demonstrated that trust in IT was not only different from trust in people, but also showed that when users have trusting beliefs in technology (recommendation agents) their perceptions about the usefulness of the technology have been positively influenced. We also argue that performance experiences would also determine their satisfaction with the OSBDS. In another study, Lankton, McKnight, and Thatcher (2014) found that the relationship with trust in technology and continuance intention could be predicted strongly by expectation disconfirmation theory, explaining 46% of the variance in continuance. Therefore, we hypothesize that:

H8. Trust in OSDST will have a positive influence on perceived usefulness of OSDST.

H9. Trust in OSDST will have a positive influence on OSDST satisfaction.

We expect perceived usefulness of OSDST to have a positive association with satisfaction. Users who believe that the technologies they are using have a positive effect on their work performance are likely to develop positive feelings about their

experiences with these technologies. Previous studies that adopt TAM and UTAUT indicate that usefulness and satisfaction are positively related. For example, TAM posits that perceived usefulness beliefs is one of the strongest determinants of attitude towards IT use behavior (Davis et al., 1999), which is supported by empirical evidence (Davis et al. 1999; Karahanna, Straub, & Chervany, 1999). In a recent study, Rezvani et al. (2017) discussed the role of perceived usefulness in the post-adoption stage, and considered usefulness as the equivalent of perceived performance. In their study that investigates the continued use for ERP systems, the authors found a significant association between perceived usefulness and satisfaction. Therefore we hypothesize that:

H10. Perceived usefulness of OSDST will have a positive influence on OSDST satisfaction.

Taking full advantage of OSDST can only be possible when individuals use them extensively, fully integrate them to their work practices, and find innovative ways of using it to perform tasks that have not even been considered before (Jaspersen et al., 2005). IS continuance model (Bhattacharjee, 2001) suggests that users are expected to continue using OSDST when they are satisfied with their use. However, post-adoption literature largely examines the relationship between satisfaction and continued use intention, which does not measure the type or depth of use behavior. From a feature-centric view of technology use, this approach is somewhat limited because it does not investigate the specific post adoption behaviors. We suggest that users who are satisfied with OSDST will have positive attitudes towards exploration behavior. In the context of our study, exploration behavior manifests itself in two

ways. One way of exploration is related to features of OSDST in use, reflected in behaviors such as using various features of OSDST, and thinking about incorporating them in work tasks (Maruping, Magni, Caporarello, & Basaglia, 2008; Thatcher et al., 2011). We also argue that users who are satisfied with OSDST will be motivated and encouraged to explore other open source technologies that might offer more effective ways of performing various work activities. Previous research has consistently found a positive link between user satisfaction and continued usage for various technologies and contexts (Bhattacharjee, 2001; Rezvani et al., 2017). We expect the influence of satisfaction to extend to exploration intentions. Similarly, post-adoption studies that examined user intent to explore found that there is a relationship between users' positive experiences and their willingness to experiment with the technology (Lippert & Forman, 2005; Thatcher et al., 2011). Therefore, we hypothesize that:

H11. OSDST satisfaction will have a positive influence on the intent to explore the OSDST in current use.

H12. OSDST satisfaction will have a positive influence on intent to explore other OSDST.

Table 7 lists all the hypothesized relationships in our proposed model.

Table 7. List of Hypotheses

No	Hypotheses
H1	Top management championship will have a positive influence on situational normality.
H2	Top management championship will have a positive influence on trust in OSDST community.
H3	Co-worker championship will have a positive influence on situational normality.
H4	Co-worker championship will have a positive influence on Trust in OSDST community.
H5	Trust in OSDST community will have a positive influence on situational normality.
H6	Trust in OSDST community will have a positive influence on Trust in OSDST.
H7	Situational normality will have a positive influence on trust in OSDST.
H8	Trust in OSDST will have a positive influence on perceived usefulness of OSDST.
H9	Trust in OSDST will have a positive influence on satisfaction.
H10	Perceived usefulness of OSDST will have a positive influence on satisfaction.
H11	Satisfaction will have a positive influence on intent to explore current OSDST.
H12	Satisfaction will have a positive influence on intent to explore other OSDST.

CHAPTER 4

METHODOLOGY

This chapter presents and describes the methods used to test the research model of this study. We will first explain the research design and the rationale behind the choice of survey methodology. Second, we will describe the sampling strategy outlining the sample frame definition as well as the sample selection process and the determination of the sample size. Third, we will introduce the research and present the definitions and measures for the constructs used in the questionnaire. Fourth, we will explain the pretest of the instrument conducted with the academic and practitioner experts in the field. Finally, we will explain our data analysis approach including the methods and tools used for the analysis of the survey data.

4.1 Research design

A well-constructed research design is considered the blueprint that provides the predetermined guidelines of an empirical study. It provides a thorough planning for the collection, measurement, and analysis of the data used to test the proposed research model (Bhattacharjee, 2012). The goal of this study is to develop and test a model to understand the factors that influence the utilization of OSDST. Since the aim is to explain the variance and develop causal relationships among the organizational, social, and technological determinants of technology utilization within organizations, we use:

- A field survey methodology for data collection purposes.
- Partial Least Squares Structural Equation Modeling (PLS-SEM) technique – to test the validity of the model and the relationships between the constructs.

Survey research is a data collection method for which written questionnaires or formal interviews are applied to gather information from a sample of respondents which then can be generalized to the rest of the population (Newsted et al., 1998). Over the years, it has emerged as the most frequently used methodology in IS research (Wang, 2015) as well as the whole of social sciences (Neuman, 2014).

Malhotra & Grover (1998) and Neuman (2014) identify the distinct characteristics of survey research:

- Data collection is conducted in a structured format, either in the form of a questionnaire or an interview.
- Data is collected from a carefully selected representative sample so the findings from the study can be generalized to the population.
- Unlike experiments, survey methodology cannot control or manipulate the conditions within a situation or context. Therefore, the use of control variables is required in order to rule out alternative explanations that could be caused by confounding variables.

The survey research process involves several stages. First, to successfully make inferences and generalize the results to the whole population, the representative sample should be identified. Then, the research instrument should be prepared, adapting established measures from previous studies whenever available. Once the instrument is ready, the next step involves the administration of the survey with the participants from the selected sample frame. Finally, after the data collection is completed, the data is analyzed to confirm the reliability and validity of the survey instrument as well as to test the hypotheses of the study.

The next sections will discuss this research process as outlined at Table 8. First, the sampling strategy will be explained. Second, the survey instrument will be

presented. Third, the approach to survey administration will be chronicled. Fourth, the choice of data analysis methods will be discussed in detail.

Table 8. Research Process of the Study

Sampling Strategy	Instrument Development	Data Collection	Data Analysis
Sample Frame	Adaptation of existing scales Pretesting	Survey administration	PLS-SEM
Sampling Method			
Sample size			

4.2 Sampling strategy

Developing a sound sampling strategy is crucial for any research as the sample chosen for the study should be representative of the population. Therefore, first and foremost, the population and sampling frame of the study are identified. Then, the chosen sampling method is justified and described in detail. Finally, the required sample size to perform data analysis with PLS-SEM analysis is calculated.

The sampling strategy of a research is closely related to the unit of analysis in a research problem (Creswell & Creswell, 2017). The unit of analysis in the study is OSDST users since the objective of the study is (a) to explore the influence of trust-building mechanisms on trusting beliefs, and (b) to investigate the impact of trusting beliefs on the utilization of OSDST.

4.2.1 Sample frame

Consistent with our research scope and unit of analysis, we identify the population of the study as the individuals who use single or multiple types of open source data science technologies in their workplaces. The sampling frame of the study consists of full-time employees in organizations that actively use OSDSTs for their work tasks.

It is not constrained by any departmental function or individual role in organizations, i.e., all employees that utilize OSDST for any data-related task that involves one or more of advanced analytics, data analysis, and data engineering activities.

4.2.2 Sampling method

Since our understanding of the availability of the sampling frame is dependent on our knowledge of individuals who use OSDST in organizations, we pursued a purposive sampling strategy. Purposive sampling is a “non-probabilistic sampling technique that allows the researchers to potentially locate all possible cases of highly specific and difficult-to-reach populations using multiple methods” (Neuman, 2014). It is highly used in field research where the cases are selected with a specific purpose, allowing researchers to pursue an in-depth investigation of a phenomenon to gain a deeper understanding of it (Neuman, 2014).

In the present study, two methods are used as part of the purposive sampling strategy. First, we contacted top management executives and senior managers from organizations where we had prior knowledge about their firm-level adoption of OSDST. Second, we conducted a LinkedIn search to identify potential OSDST users in Turkey. Our search revealed around 5000 individuals in Turkey that included “analytics”, “data science” and/or “big data” within their job titles or descriptions. We randomly selected these individuals from the list and checked if they are currently employed in an organization. Next, we contacted the selected individuals asking for their participation in the study if they were using one or more of the OSDST as part of their jobs. If the answer was positive, we included those individuals in our sample set. If not, we eliminated those individuals. Since reaching the desired sample size was a foreseen challenge, we considered this method and

LinkedIn as a valuable source in case the direct data collection efforts from organizations did not result in an adequate number of responses.

4.2.3 Sample size

Because of the novelty of the organizational diffusion of OSDST in organizations as well as the difficulties related to the identification of employees who actively use these technologies, obtaining a large sample has been challenging. Our final sample size for data analysis is 173, which meets three criteria recommended and commonly cited in the literature.

While co-variance-based SEM approach is relatively strict in required sample size to ensure adequate statistical power and test hypothesized relationships, PLS-SEM provides researchers the flexibility to work efficiently with small sample sizes and complex models since it employs OLS regressions to estimate partial regression relationships instead of doing it for the whole model at the same time (Cassel, Hackl, & Westlund, 1999; Hair et al., 2017). As reported by Hair et al. (2017), many studies have demonstrated that PLS-SEM displays high levels of statistical power with small sample sizes even with complex model structures (e.g., Reinartz, Haenlein, & Henseler, 2009), although it is also acknowledged that larger sample sizes are beneficial to increasing the precision of PLS-SEM estimations (Hair et al., 2017).

While PLS-SEM performs well with small sample sizes, there are several methods researchers can apply to calculate the necessary sample size that provides adequate statistical power and reliable path coefficient estimations. One benchmark that is the widely used is Barclay et al.'s (1995) "10-times rule", which recommends that the sample size should at least be 10 times the largest number of indicators used to measure a particular construct in the structural model. In the present study, the

highest number of indicators to measure a first-order construct is 4. Therefore, according to this method, a minimum sample size of 40 is deemed sufficient.

Another recommended method for PLS-SEM sample size is developed by Cohen (1992). Cohen's method considers not only the number of construct indicators but also the model characteristics. According to the guideline of this method, for a statistical power of 80%, at the desired statistical significance level of $\alpha=0.05$, and to estimate a minimum R^2 of 0.10, the sample size should at least be 113.

As the third option, we employed the inverse square root method (that is developed and recommended by Kock and Hadaya (2018) which suggest that "the probability that ratio of a path coefficient and its standard error will be greater than the critical value of a test statistic for a significance level". According to the inverse square root method, assuming a power level of 80%, a statistical significance level of $\alpha=0.05$, and a minimum path coefficient of 0.2, the minimum sample size recommended is 155.

4.3 Survey instrument development

Previously developed, validated, and widely used scales are adapted to measure the constructs in the study. Since data science technologies and their use in organizations are relatively recent, preexisting items were refined based on relevant literature review, practitioner white papers as well as interviews with data science professionals. Next, we provide the definition and operationalization of the constructs used to test the hypothesized relationships of the study. In order to eliminate potential alternative explanations, personal factors such age, gender, experience and industry are used as control variables.

In the opening section of the research instrument, we defined open source data science and advanced analytics technologies as:

1. Next Generation Mathematical Modeling and Machine Learning Platforms and Programming Languages: They refer to the open source technologies that enable end-to-end automation of all analytical activities (e.g., R, Python, Spark/Scala, TensorFlow, PyTorch, Julia).
2. Advanced Data Management Technologies: They refer to the open source technologies used to build highly scalable data systems. (e.g., Hadoop, NoSQL Databases, Stream Processing Engines)

4.3.1 Top management championship

Top management championship of OSDST refers to managerial beliefs about the open source data science initiatives and management's active participation in these initiatives. Taken together, "top management belief" and "top management participation" are the two reflective indicators (second-order constructs) of top management championship (first-order construct). Both scales are adapted from Chatterjee et al. (2002), and Liang, Saraf, Hu, and Xue (2007).

Top management belief measures reflect the extent to how the technology users perceive their senior management's belief in the operational and strategic benefits of OSDST in terms of how using them can improve business activities as well as create competitive advantage over other firms in the market.

- Stem: Think about the senior management of your firm and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

TCB1: The senior management of our firm believes that open source data science technologies have the potential to provide significant business benefits to the firm.

TCB2: The senior management of our firm believes that open source data science technologies will create a significant competitive advantage for firms.

TCB3: The senior management of our firm believes that it is not necessary to use open source data science technologies to conduct business activities.

On the other hand, top management participation scale measures the extent to how the technology users perceive their senior management's active participation in the adoption of OSDST via articulating a vision, formulating a strategy, and establishing goals and standards for the use of OSDST within an organization (Chatterjee et al., 2002).

- Stem: Think about the senior management of your firm and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

TCP1: The senior management of our firm actively articulates a vision for the organizational use of open source data science technologies.

TCP2: The senior management of our firm actively formulated a strategy for the organizational use of open source data science technologies

TCP3: The senior management of our firm actively established goals and standards to monitor the open source data science technologies projects.

4.3.2 Co-worker championship

Co-worker championship refers to co-workers' beliefs about open source data science technologies and their active participation in those technologies. Like top management championship, co-worker championship also has two reflective indicators that are "Peer Belief" and "Peer Participation". Again, the scales were adapted from Chatterjee et al. (2002) and Liang et al. (2007). Co-worker belief items measure the extent to how the technology users perceive their co-workers' belief in the operational and strategic benefits of OSDST in terms of how using them can improve operational activities as well as creating business advantages to their departments. On the other hand, co-worker participation scale measures how the technology users perceive their co-workers' active participation via their use of OSDST for data manipulation and analysis activities as well as in in-production projects.

- Stem: Think about the co-workers in your department and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

PCB1: My co-workers believe that open source data science technologies have the potential to provide significant business benefits to our department.

PCB2: My co-workers believe that open source data science technologies will create a significant operational advantage for our department.

PCB3: My co-workers believe that it is not necessary to use open source data science technologies to conduct our departmental activities.

- Stem: Think about the co-workers in your department and indicate the extent to which you agree or disagree with these statements.

- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

PCP1: My co-workers actively learn different open source data science technologies to enhance their data manipulation and/or analysis skills.

PCP2: My co-workers actively experiment with open source data science technologies during the project development and testing processes.

PCP3: My co-workers actively use open source data science technologies for in-production projects.

4.3.3 Trust in OSDST community

Since there was no pre-existing scale of general trust in communities and online OSDST community, the items developed for this study are derived from McKnight et al (2002), with at least one item that measures the reliability, functionality, and helpfulness perceptions about the OSDST community in general.

- Stem: Think about data science communities in general and indicate the extent to which you agree or disagree with these statements.

- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

TCOM1: Open source data science communities have enough resources to make me feel comfortable to use these technologies.

TCOM2: I feel confident that I will find reliable solutions in OSDS communities if needed.

TCOM3: I am comfortable finding information that I need on OSDS communities.

TCOM4: If I needed help, I feel confident that someone from the community would help me.

4.3.4 Situational normality

It refers to the belief that success with open source data science technologies is likely because an individual feels normal and comfortable using them at the workplace. The items are derived from McKnight et al. (2011), who adapted the measures from their earlier work (McKnight et al., 2002) into trust in technology context. The reliability of the construct was reported as 0.94.

- Stem: Think about the open source data science technologies in general and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

TSN1: I am totally comfortable working with open source data science technologies.

TSN2: I feel very good about how things go when I use open source data science technologies.

TSN3: I always feel confident that the right things will happen when I use open source data science technologies.

TSN4: It appears that things will be fine when I utilize open source data science technologies.

4.3.5 Trust in OSDST

Like top management championship and co-worker championship, trust in OSDST is also a second-order reflective construct comprised of three first order constructs that are “functionality”, “helpfulness”, and “predictability” (McKnight et al., 2011). All items are adapted from McKnight et al. (2011) and Thatcher et al. (2011), In both studies, reliabilities for all three first order constructs are reported to have a reliability over 0.90.

Functionality belief refers to the belief that the OSDST have the capability, functions, or features for the users to perform the tasks.

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

TTF1: Open source data science technologies that I am currently using have the functionality that I need.

TTF2: Open source data science technologies that I am currently using have the features required for my work activities.

TTF3: Open source data science technologies that I am currently using have the overall capabilities I need.

Helpfulness belief refers to the belief that the OSDST provide adequate and responsive aid.

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.
- Scale: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

TTH1: The help function of the open source data science technologies that I am currently using provides competent guidance.

TTH2: The help function of the open source data science technologies that I am currently using provides whatever help I need.

TTH3: The help function of the open source data science technologies that I am currently using provides very sensible and effective advice if needed.

TTH4: The help function of the open source data science technologies that I am currently using supplies my need for help through a help function.

Predictability belief refers to the belief that the OSDST act consistently and its behavior can be forecast.

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

TTP1: The open source data science technologies that I am currently using behaves in a predictable way.

TTP2: I can forecast in advance how the open source data science technologies that I am currently using will work.

TTP3: The open source data science technologies that I am currently using functions in the same way each time I use them.

TTP4: As a work tool, the open source data science technologies that I am currently using are very predictable.

4.3.6 Perceived usefulness

It refers to the degree to which a user believes that using open source data science technologies enhance his or her job performance. The scale is adapted from Davis (1989) and was reported to have high reliability of 0.95 (Thatcher et al., 2011) in a study that examined trust in technology in a knowledge management system context.

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)
- Items:

PU1: Using open source data science technologies improves my performance at work.

PU2: Using open source data science technologies enhances my effectiveness in completing work tasks.

PU3: Using open source data science technologies enhances my productivity in completing projects.

4.3.7 Satisfaction

It refers to a user's feelings about the open source data science technologies that he/she has experienced. The items are derived from Bhattacharjee (2001), which was adapted from Spreng, MacKenzie, and Olshavsky's (1996) overall satisfaction scale. Bhattacharjee (2001) reported that the 4-item scale had a reliability of 0.87.

- Stem: Think about your overall experience with the open source data science technologies you have used so far and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

SAT1: Very dissatisfied/Very satisfied.

SAT2: Very displeased/Very pleased.

SAT3: Very frustrated/Very contended.

SAT 4: Absolutely terrible/Absolutely delighted

4.3.8 Intent to explore

Previous literature has developed and tested measures for user intention to explore specific technologies. These studies had the purpose of understanding postadoption behavior of infusing technologies under consideration when using them for tasks.

The items are adapted from Nambisan et al. (1999) and Thatcher et al. (2011), who reported acceptable reliability for the construct in previous research (both over 0.90)

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.

- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

CEXP1: I intend to spend considerable time exploring uses of open source data science technologies that I am currently using.

CEXP2: I intend to explore the open source data science technologies I am currently using for enhancing the effectiveness at work.

CEXP3: I intend to use the open source data science technologies I am currently using in novel ways.

Intent to explore technologies in use measures how the individuals are willing to utilize them further. On the other hand, intent to explore other OSDST measures their intent to investigate and experiment with technologies that the individuals

currently not using but may find beneficial for discovering new ways of working with data or improving their skill sets. The measures for this construct are also derived from Nambisan et al. (1999) and Thatcher et al. (2011).

- Stem: Think about the open source data science technologies you are using and indicate the extent to which you agree or disagree with these statements.
- Scale Range: (Strongly Disagree = 1; Strongly Agree = 7)

- Items:

GEXP1: I experiment with new open source data science technologies for potential ways of analyzing data.

GEXP2: I investigate new open source data science technologies for enhancing my ability to perform manipulating and/or analyzing data.

GEXP3: I spend considerable time in exploring new open source data science technologies to help me manipulate and/or analyze data.

GEXP4: I invest substantial effort in exploring new open source data science technologies.

4.3.9 Instrument pretesting

After completing the survey instrument and preparing an online version of the questionnaire, and before administering the survey, we conducted a pretest.

Pretesting is the preliminary examination of the research instrument to identify potential problems that might lead to unexpected difficulties or problems for the participants (Alreck & Settle, 1995; Presser et al., 2004). Other researchers also point out the importance of field-based pretesting and recommend the pretesting of the instrument with the expert practitioners in the field. This is done in order to ensure

that the operationalizations of the constructs reflect the real world (Malhotra & Grover, 1998). Accordingly, the aim of pretesting the instrument is:

- To ensure that the conceptualization of the problem match the actual experiences of practitioners for the appropriate operationalization of the items (content validity)
- To ensure that the Turkish translation of the survey items are accurate and precise as most of the items had been originally developed and validated in English.
- To fix any problems or errors that the pretest participants would detect or to make improvements as they would suggest.

We conducted the pretest with the participation of three academic experts in Management and Information Systems fields, sixteen Management PhD students at Boğazici University, and eight practitioners of the data science/advanced analytics. The representative industries of practitioners consisted of telecommunications, banking, retail, and technology.

An online version of the survey was sent to the practitioners via email while a paper version was shared with the faculty members and PhD students. All participants provided feedback both via interviews (15-30 minutes) and shared written comments about the questionnaire.

Overall, the pretest results revealed two problem areas:

- Despite our meticulous efforts for the translation, participants raised concerns about some of the Turkish wordings/translations of some items.
- Some of the organization-related items were found to be sensitive. In addition to warranting anonymity, participants suggested leaving such questions to the end of the survey questionnaire.

Following these suggestions and after several iterations, the instrument was finalized for the data collection phase.

4.4 Survey administration

In order to reach as many respondents as possible and expedite the data collection process, survey questionnaire is administered digitally using the Google Forms platform online. Online surveys have been ubiquitous in recent academic research, providing advantages such as lower cost, increased reach, flexibility, speed, and timeliness (Evans & Mathur, 2018).

After respondents agreed to participate in the study, we have provided them the survey link via email or LinkedIn depending on their preferences. The introduction page of the questionnaire (Appendix B) included the research objective and clarified the sampling frame to ensure capturing the most representative sample. We have also notified the participants that their participation was voluntary and there were no monetary incentives offered by the researcher. Because the survey instrument consists of sensitive questions regarding an individual's perceptions about his/her top management and co-workers, we ensured the participants that their responses were collected anonymously, and all data would be analyzed after aggregation. Also, in line with the expert recommendations at the pretest phase, we did not ask for any personal information that a participant would feel that can be tracked back to her. Since the survey was not administered in a physical environment, we have provided our contact information in case the participants may have any questions or concerns.

The questions were ordered carefully with the consideration of recommendations at the pretest stage of research. The questionnaire started with less

sensitive questions, making sure that they complete most of the survey before they reach the organizational questions. Even then, we created stems that reminded the participants about the anonymity of the responses and motivated them with importance of their participation. We did not set any of the questions mandatory to avoid dropouts in case a participant wishes to not respond to a question.

First, we asked the respondents open-ended questions regarding the use of open source data science technologies in organizations. They included their personal experience with OSDST (in duration) and the OSDST that they actively use at their workplace. In addition to understanding the most widely used technologies in organizations, these questions are utilized also for data screening purposes, since they confirm their usage. Accordingly, any cases with blank responses for these two questions are excluded from the data analysis during the data preparation stage. At the end of the survey, demographic questions that are used as control variables are introduced:

4.5 Data analysis procedure

We selected Partial Least Squares Structural Equation Modeling (PLS-SEM) as the chosen data analysis methodology where we both validate our construct measures (measurement model) and test the hypotheses of the research model (structural model).

PLS-SEM is one of the two major approaches of Structural Equation Modeling (SEM), which has become the de facto methodology for survey studies that use latent variables and examine complex relationships between the constructs. Before the emergence of SEM, previous studies that used latent variables and examined multiple dependent variables relied on first-generation regression models that had the ability to test the effects on dependent variables only in independent analyses, while

the construct validity could be assessed separately using exploratory or confirmatory factor analysis. SEM, considered a second generation regression technique, allows researchers to conduct the factor analysis and test the hypothesized relationships with multiple independent and dependent variables simultaneously in the same analysis (Gerbing & Anderson, 1998; Gefen, Straub, & Boudreau, 2000).

SEM approaches can be classified as covariance based SEM (CB-SEM), which is based only on common variance, and PLS-SEM which is based on total explained variance (Hair et al., 2016). While CB-SEM treats constructs as common factors whose scores are not known or needed in the estimation of model parameters, PLS-SEM relaxes the assumptions of CB-SEM with its composite-based approach where the covariation among the indicators is explained by a common factor (Hair et al., 2017) The main differences between the two approaches are summarized in Table 9.

PLS-SEM's capabilities that allow researchers to work with small size samples, non-normal data, and complex models even with formative constructs contributed to wide application of this method in the Information Systems field (e.g., Gefen et al., 2011). PLS-SEM is also suitable for the present study for the following two reasons:

- Exploratory aspect of the study: Hair et al. (2017) suggest that exploratory research can be used to test hypotheses in quantitative research especially when researchers face problems that cannot be well-defined, or researchers cannot have a comprehensive knowledge to propose hypotheses to explain all relationships. Although we developed our research model based on established theories, the use of OSDST in organizations, especially for individuals in business units, is a recent issue that has not been investigated before. Therefore, one of our objectives is to understand and explain any

Table 9. PLS-SEM vs CB-SEM

Issue	PLS-SEM	CB-SEM	Reference
Sample size	Works well with relatively small sample size while maintaining statistical power	Large sample size is required to estimate model parameters	Chin and Newsted, 1999 Reinartz et al., 2009
Non-normal data	Non-parametric approach relaxes the multivariate assumption that allows working with non-normal data.	Parametric approach with strong multivariate assumptions that require normal data	Hair et al., 2017
Model complexity	Both reflective and formative constructs can be assessed in the same model	Cannot estimate models that include both reflected and formative constructs	Albers, 2010 Hair et al., 2017
Model confirmation	Reliability Validity	Reliability Validity Goodness of fit	Hair et al., 2017
Purpose	Both exploratory and confirmatory research	Mainly for confirmatory research	Hair et al., 2017

contradictory findings as well as examine potential moderations during the post-hoc analysis of the study.

- Small sample size: As explained earlier, due to the difficulties involved in identifying and locating representative individuals in the sampling frame, our final sample size is 173. Because PLS-SEM uses ordinary least squares (OLS) regression which estimates the path coefficients as separate subparts, the complexity of the model does not have much influence on sample size requirements (Hair et al., 2017).

CHAPTER 5

DATA ANALYSIS AND RESULTS

In this chapter, we will first explain the data screening and preparation process. After preparing the data for analysis, we will present the sample characteristics. Then the research model will be tested employing the PLS-SEM methodology using the SmartPLS software version 3.3.9. The measurement model results will be reported to demonstrate the reliability and construct validity of the measures, followed by the findings of the structural model where the hypothesized relationships are tested. Finally, we will investigate the differences between the categorical control variables, namely departmental units, gender, age.

5.1 Data screening

We screened the data first to ensure the quality of the data and the suitability of the analysis techniques. First, we examined the survey engagement of the respondents using the standard deviation of the answers. We checked if any of the cases demonstrated standard deviation below 0.5 to assess respondent engagement. 1 case was found and deleted due to low engagement in answers. 15 more cases were deleted since respondents have stated that either they and/or their organizations are not actively using open source data science technologies yet. Out of 192 respondents, 3 cases are deleted to missing values over 15%. Missing values for the variables have not exceeded 2% of the responses. The missing data was treated with mean imputation method before running the PLS-SEM analysis. After cleaning the data, the final data set for analysis had a sample size of 173.

The normality of the data distribution was also assessed. Although PLS-SEM's non-parametric approach does not require normality, extremely non-normal data is known to threaten the precision of the statistical tests performed by PLS Bootstrapping Algorithm by distorting the standard errors (Hair et al., 2017). As reported in Appendix D, we examined the skewness and kurtosis value of each item used in the present study. According to Hair, Black, Babin, and Anderson (2010) data is regarded as normal when skewness values lie between - 2 and +2 and kurtosis lies between - 7 and +7. Our results reveal that our data is largely normal, with one item, PU3 has a kurtosis of 7.119. Therefore, there is no sign of an extreme non-normality in the data that would impact the hypotheses testing of the structural model.

5.2 Sample characteristics

We asked the respondents some of their demographic properties as part of the survey. We used this demographic information (age, gender, industry) as the control variables of the study as well as for further investigation of potential moderating relationships. We also asked the respondents the time lapsed after their initial adoption to gain more information about their experience with OSDST. In order to understand the choice of OSDST in organizations, we also asked which open source data science technologies they are actively using at work.

The most frequent age range of the respondents is 25-34 (62.4%), followed by 35-44 (27.2%). Only 1% of the respondents fall between the ages of 45 and 54. This may indicate that individuals who use OSDST are relatively younger employees at workplaces.

Table 10. Respondents by Age

Age	Frequency	Relative Frequency
18-24	16	9.25%
25-34	108	62.43%
35-44	47	27.17%
45-54	2	1.16%
Total	173	100.00%

Table 11 shows that the sample consists of mostly male technology users. 64.1% of the OSDST are male while 35.3% of them are female. This may suggest that OSDST are adopted more by men compared to women.

Table 11. Respondents by Gender

Gender	Frequency	Relative Frequency
Male	111	64.16%
Female	61	35.26%
Missing	1	0.58%
Total	173	100.00%

While all respondents hold at least an undergraduate degree as expected, almost half of them have graduate degrees (47.4%), including 2.3% with a PhD degree. This is perhaps due to the complexity and difficulty of developing the skills that require a certain level of understanding in various areas such as computer science, statistics, and machine learning. The open source technologies in data science may be utilized by highly educated individuals because working as a data scientist/advanced analytics practitioner demands advanced knowledge within these domains.

Table 12. Respondents by Education

Education	Frequency	Relative Frequency
Undergraduate	86	50.00%
Graduate	82	47.40%
PhD	4	2.31%
Missing	1	0.58%
Total	173	100.00%

Most commonly used OSDST is Python, followed by R. Although respondents report data storage and engineering technologies as well, the findings indicate that most of the users in the sample are using open source programming languages for their data-related tasks at work.

Table 13. OSDST Used in Organizations

Technology	Frequency
Python	124
R	47
Spark	26
NoSQL	10
Hadoop	16
Scala	7
PyTorch	6
Postgre SQL	5

More than half our respondents are from the banking industry (57.2%). In addition, technology and telecom companies amount to more than 10% of the sample each. Although more than 10 other industries are represented, they constitute roughly 17% of the whole sample.

Table 14. Respondents by Industry

Industry	Frequency	Relative Frequency
Banking	99	57.23%
Technology	24	13.87%
Telecom	21	12.14%
E-Business	6	3.47%
Retail	4	2.31%
Manufacturing	4	2.31%
Holding	3	1.73%
Education	3	1.73%
Insurance	2	1.16%
Media	1	0.58%
Health	1	0.58%
Defense	1	0.58%
Agriculture	1	0.58%
Other	3	1.73%
Total	173	100.00%

5.3 Measurement model

This section reports the evaluation of the measurement model in terms of its reliability and construct validity, which consists of convergent validity and discriminant validity. The summary of the results for all reliability and validity assessments are reported in Table 18.

5.3.1 Reliability

Assessing the measurement model involves ensuring the reliability and the construct validity of the instrument scales used in the questionnaire. Reliability refers to the extent to which the variables used in the instrument is consistent in what they are intended to measure. We test reliability using three established techniques (Cronbach's α , composite reliability, and Rho_A) that complement each other. The values of all three techniques range between 0 and 1, with the expectation that higher values would demonstrate higher reliability among the variables.

The most traditional method of reliability is Cronbach's α , for which the recommended minimum values suggested are .60 (Nunnally, 1967) or .70 (Nunnally and Bernstein, 1994) for each construct. Cronbach's α for each construct in the present study is over .70, with values that range from 0.818 (TTP) to 0.915 (MCP). Because Cronbach's α is suspected to be sensitive to the number of items used to measure a construct and it can underestimate the internal consistency reliability scores, Hair et al. (2017) recommend the additional use of composite reliability, with recommended minimum values of 0.70 or 0.90. All constructs indicate high composite reliability, ranging from 0.879 (TSN) to 0.947 (MCP).

Table 15. Construct Reliability and Convergent Validity

	Cronbach's Alpha	rho_A	Composite Reliability	AVE
MCB	0.875	0.875	0.941	0.889
MCP	0.915	0.917	0.947	0.855
PCB	0.846	0.846	0.928	0.867
PCP	0.903	0.905	0.939	0.838
TSN	0.819	0.839	0.879	0.645
TCOM	0.902	0.906	0.932	0.774
TTP	0.818	0.823	0.891	0.733
TTF	0.866	0.866	0.918	0.789
TTH	0.835	0.873	0.891	0.675
PU	0.882	0.888	0.927	0.809
SAT	0.882	0.915	0.917	0.735
GEXP	0.864	0.886	0.916	0.785
CEXP	0.897	0.904	0.929	0.766

The third technique, Rho_A, is very similar to composite reliability in terms of assessment, with the only difference being the use of unstandardized outer loadings of the indicator variables. Again, all constructs demonstrate high Rho_A scores, ranging from 0.823 (TTP) to 0.917 (MCP). Overall, these results provide support for the reliability of the constructs in the research instrument.

5.3.2 Convergent validity

While reliability indicates the consistency of our measures, construct validity indicates the extent to which the operationalization of a construct measures the concepts that it intends to measure (Straub, 1989). In this regard, first we examine how well a particular construct measures what it intends to measure. This type of construct validity is called convergent validity, for which we expect the items suggested to reflect a construct to converge together (Straub, Boudreau, & Gefen, 2004). Hence, we expect the items to show high correlations with one another, particularly when compared to the convergence of items relevant to other constructs (Straub et al., 2004).

For convergent validity, we look at the outer loadings at Appendix E, the average variance extracted (AVE), and indicator reliability for each construct. The initial results showed that the outer loadings of several items (MCB3, PCB2, TTP3) were significantly below the recommended minimum of 0.70 (Hair et al., 2017), influencing the indicator reliability and AVE values negatively as well. Since there are studies that suggest a minimum level of .60 is also acceptable (e.g., Chin, 1998) we decided to keep TTH4 (0.625). After the removal of the three items, the measurement model was tested again, this time meeting the convergent validity requirements. Indicator reliability values of all items except one (TTH4, 0.390) were

found to be over the recommended 0.50 threshold (Hair et al., 2017). To assess convergent validity, AVE values for each construct should be larger than its correlation with other constructs, and each construct should meet the recommended minimum criterion of 0.50. (Fornell & Larcker, 1981). The AVE of every construct is above 0.50, ranging from 0.645 (TSN) to 0.889 (MCB). Therefore, the convergent validity of the measurement model is supported.

5.3.3 Discriminant validity

After the convergent validity of the measures is confirmed, we test the second component of construct validity, discriminant validity. While convergent validity shows how well a group of items converge together to reflect or form a construct, discriminant validity assures that these items measure only what they are intended to measure and not any of the other constructs. There are several approaches to verifying discriminant validity. First, the factor loadings of the items on their theoretically assigned constructs should be higher than their loadings on any other construct in the model (Gefen & Straub, 2005). Cross loadings presented at Appendix F demonstrates that none of items load higher on any other construct. Second, we examine the heterotrait-monotrait ratios (HTMT) of the constructs at Table 16. According to Hair et al. (2017), the confidence interval of the constructs should not include 1.00. No construct in our measurement model includes 1.00 within their confidence intervals. Our final approach to assessing discriminant validity is the AVE analysis based on the Fornell-Larcker criterion. According to the criterion, the square roots of AVEs for each construct should be larger than its correlation with any of the constructs in the measurement model (Gefen & Straub,

2005). Table 17 demonstrates that the Fornell-Larcker criterion is met, indicating that the discriminant validity our measurement model is also supported.

Table 16. Discriminant Validity: Hetero-Monotrait Ratio (HTMT)

	MCB	MCP	PCB	PCP	TSN	TCOM	TTP	TTF	TTH	PU	SAT	GEXP	CEXP
MCB													
MCP	0.857												
PCB	0.628	0.541											
PCP	0.500	0.489	0.906										
TSN	0.131	0.150	0.272	0.217									
TCOM	0.166	0.136	0.306	0.159	0.413								
TTP	0.085	0.133	0.257	0.193	0.626	0.533							
TTF	0.099	0.085	0.236	0.153	0.603	0.407	0.580						
TTH	0.224	0.270	0.263	0.225	0.340	0.424	0.474	0.477					
PU	0.093	0.068	0.293	0.220	0.625	0.270	0.449	0.500	0.189				
SAT	0.094	0.125	0.205	0.148	0.623	0.445	0.634	0.848	0.467	0.584			
GEXP	0.065	0.132	0.099	0.079	0.398	0.125	0.227	0.288	0.222	0.271	0.302		
CEXP	0.138	0.172	0.148	0.089	0.474	0.179	0.281	0.388	0.102	0.248	0.380	0.729	

Table 17. Discriminant Validity: Fornell-Larcker Criterion

	MCB	MCP	PCB	PCP	TSN	TCOM	TTP	TTF	TTH	PU	SAT	GEXP	CEXP
MCB	0.943												
MCP	0.768	0.925											
PCB	0.540	0.443	0.931										
PCP	0.443	0.475	0.791	0.916									
TSN	0.116	0.131	0.234	0.198	0.803								
TCOM	0.148	0.125	0.267	0.143	0.375	0.880							
TTP	0.072	0.112	0.213	0.165	0.529	0.454	0.856						
TTF	0.087	0.070	0.202	0.136	0.516	0.360	0.493	0.888					
TTH	0.191	0.231	0.226	0.208	0.296	0.359	0.399	0.424	0.822				
PU	0.080	0.048	0.253	0.196	0.540	0.242	0.384	0.437	0.150	0.900			
SAT	0.084	0.102	0.178	0.132	0.551	0.401	0.548	0.750	0.420	0.526	0.875		
GEXP	0.043	0.114	0.089	0.074	0.356	0.119	0.210	0.261	0.200	0.250	0.278	0.857	
CEXP	0.123	0.160	0.129	0.064	0.398	0.159	0.242	0.342	0.088	0.216	0.340	0.609	0.886

Table 18. Reliability and Validity: Summary

Latent Variable	Indicators	Convergent Validity			Internal Consistency Reliability		Discriminant Validity
		Loadings	Indicator Reliability	AVE	Composite Reliability	Cronbach's Alpha	
		>0.70	>0.50	>0.50	0.60-0.90	0.60-0.90	HTMT Confidence Interval Not Including 1
MCB	MCB1	0.942	0.887	0.889	0.941	0.875	Yes
	MCB2	0.944	0.891				
MCP	MCP1	0.926	0.857	0.855	0.947	0.915	Yes
	MCP2	0.950	0.903				
	MCP3	0.898	0.806				
PCB	PCB1	0.930	0.866	0.867	0.928	0.846	Yes
	PCB3	0.931	0.867				
PCP	PCP1	0.913	0.834	0.838	0.939	0.903	Yes
	PCP2	0.949	0.901				
	PCP3	0.883	0.780				
TSN	TSN1	0.809	0.654	0.645	0.879	0.819	Yes
	TSN2	0.795	0.633				
	TSN3	0.850	0.723				
	TSN4	0.755	0.570				
TCOM	TCOM1	0.858	0.736	0.774	0.932	0.902	Yes
	TCOM2	0.916	0.839				
	TCOM3	0.915	0.837				
	TCOM4	0.826	0.682				
TTP	TTP1	0.846	0.715	0.733	0.891	0.818	Yes

	TTP2	0.839	0.704				
	TTP4	0.883	0.779				
TTF	TTF1	0.871	0.758	0.789	0.918	0.866	Yes
	TTF2	0.924	0.853				
	TTF3	0.869	0.755				
TTH	TTH1	0.897	0.805	0.675	0.891	0.835	Yes
	TTH2	0.895	0.801				
	TTH3	0.839	0.703				
	TTH4	0.625	0.390				
PU	PU1	0.885	0.784	0.809	0.927	0.882	Yes
	PU2	0.904	0.816				
	PU3	0.910	0.827				
SAT	SAT1	0.906	0.821	0.735	0.917	0.882	Yes
	SAT2	0.912	0.831				
	SAT3	0.894	0.798				
	SAT4	0.782	0.612				
GEXP	GEXP1	0.888	0.789	0.785	0.916	0.864	Yes
	GEXP2	0.887	0.786				
	GEXP3	0.855	0.731				
	GEXP4	0.796	0.633				
CEXP	CEXP1	0.856	0.733	0.766	0.929	0.897	Yes
	CEXP2	0.929	0.863				
	CEXP3	0.872	0.760				

5.3.4 Multicollinearity and common method bias

Before testing the hypotheses of the structural model, we check if the measures suffer from any multicollinearity issues or suffer from common method bias. PLS-SEM allows us to assess both potential problems with the use of the variance inflation factor (VIF). While the literature recommends several thresholds, they vary greatly. We refer to the recent guidelines and one particularly recommended for SEM-based research. Kline (1998) and Kock (2015) both recommend VIF values below 5.0. As reported in Appendix G, outer VIF values for each item range from 1.329 (TTH4) to 4.784 (MCP2). Therefore, we conclude that there is no significant threat of multicollinearity or common method bias in the research instrument.

5.4 Structural model

5.4.1 Hypotheses testing

In order to test the structural model and estimate the path coefficients between the constructs, the PLS Bootstrapping procedure is applied. As recommended by the SmartPLS, a bias-corrected bootstrapping is applied at a significance level of 0.05 to acquire 5000 resampled cases. To interpret the structural model, we examine the path coefficients (similar to standardized betas in OLS-regression) and test their t-values against a set significant level of α . While p-values display the significance of hypothesized paths, R^2 values indicate the explained variance of endogenous variables. To estimate the substantiality of the effect sizes, f^2 values of the relationships between the statistically significant relationships are also examined.

Findings of the PLS Bootstrapping procedure indicate that most of the path coefficients are significant (Table 19). Overall, most of the hypotheses of the research model are supported (Table 20). Our first set of hypotheses investigates the

relationships between organizational factors and general trusting beliefs. H1 and H2 hypothesize top management championship, one of the organizational factors, to have a positive influence on trust in OSDST community as well as situational normality. However, neither of the hypotheses is statistically significant at any acceptable level. H3 and H4 hypothesize co-worker championship to have a positive influence on both constructs. The relationship between co-worker championship and trust in OSDST community ($p=0.077$) and co-worker championship and situational normality ($p=0.097$) are significant only at $\alpha=0.1$ level, demonstrating weak support for H3 and H4.

Table 19. Path Coefficients and Statistical Significance

	O.S.	S.M..	S.D.	t-stat	p-value
MGMT Champ -> Trust in- OSDST Community	0.050	0.052	0.099	0.501	0.617
MGMT Champ -> Situational Normality	0.005	0.011	0.106	0.043	0.966
Co-worker Champ -> Trust in OSDST Community	0.179	0.187	0.101	1.770	0.077
Co-worker Champ -> Situational Normality	0.152	0.156	0.092	1.662	0.097
Trust in OSDST Community -> Situational Normality	0.343	0.341	0.082	4.164	0.000
Trust in OSDST Community -> Trust in OSDST	0.324	0.326	0.064	5.067	0.000
Situational Normality -> Trust in OSDST	0.446	0.449	0.055	8.053	0.000
Trust in OSDST -> Usefulness	0.415	0.419	0.068	6.095	0.000
Trust in OSDST -> Satisfaction	0.622	0.624	0.049	12.566	0.000
Usefulness -> Satisfaction	0.268	0.269	0.067	3.995	0.000
Satisfaction -> Intent to Explore (Others)	0.278	0.289	0.069	4.044	0.000
Satisfaction -> Intent to Utilize (Current)	0.339	0.349	0.066	5.118	0.000

o.s.=original sample, s.m.=sample mean, s.d.=standard deviation

Next, we examine the relationships between the general trusting beliefs as well as their influence on users' trust in OSDST they currently use as part of their jobs. The relationship between trust in OSDST community and situational normality lends support for H5 ($p < 0.001$). Results also indicate strong support for H6 ($p < 0.001$) which suggest trust in OSDST community to influence trust in OSDST positively. H7 is also strongly supported ($p < 0.001$), providing evidence that situational normality is significantly related to trust in OSDST. Together, these factors contribute to more than 16% of the explained variance in situational normality with OSDST ($R^2 = 0.163$) although they accounted only for 4% of the explained variance in Trust in OSDST community ($R^2 = 0.044$).

We hypothesize a positive relationship between general and specific trusting beliefs (H6-H7). Findings demonstrate significant relationships between trust in OSDST community and trust in OSDST ($p < 0.001$) as well as situational normality and trust in OSDST ($p < 0.001$), providing support for both hypotheses. The model explained more than 40% of the variance in trust in OSDST ($R^2 = 0.413$).

As part of the post adoption hypotheses, we posit that trust in OSDST would have a positive impact on users' perceptions of the usefulness of these technologies (H8) and their satisfaction with them (H9). There is strong support for both hypotheses ($p < 0.001$), indicating the importance of trust in OSDST during the continued use phase. H10 hypothesizes a positive relationship between perceived usefulness of OSDST and satisfaction with OSDST and is supported strongly as well ($p < 0.001$). Overall, the model relationships contribute to nearly 60% of the explained variance in users' satisfaction with OSDST.

Finally, we test if satisfaction with OSDST would lead to the utilization and exploration of these technologies further. Hence, we hypothesize a positive

relationship between satisfaction and intent to explore the OSDST that the users currently use (H10) as well as satisfaction and intent to explore other OSDST (H11). Both hypotheses are supported strongly ($p < 0.001$).

After testing the statistical significance of the proposed relationships, we examine the effect sizes of the significant relationships. In PLS-SEM research, an f^2 value over 0.02 is considered small, over 0.15 is considered medium, and over 0.35 is considered large effect sizes (Cohen, 1992; Kock, 2014). The effect sizes of the present study are reported in Table 21.

Table 20. Hypotheses Testing Results

No	Hypotheses	Support
H1	Top management championship will have a positive influence on situational normality.	No
H2	Top management championship will have a positive influence on trust in OSDST community.	No
H3	Co-worker championship will have a positive influence on situational normality.	Yes*
H4	Co-worker championship will have a positive influence on Trust in OSDST community.	Yes*
H5	Trust in OSDST community will have a positive influence on situational normality.	Yes
H6	Trust in OSDST Community will have a positive influence on Trust in OSDST.	Yes
H7	Situational normality will have a positive influence on trust in OSDST.	Yes
H8	Trust in OSDST will have a positive influence on perceived usefulness of OSDST.	Yes
H9	Trust in OSDST will have a positive influence on satisfaction.	Yes
H10	Perceived usefulness of OSDST will have a positive influence on satisfaction.	Yes
H11	Satisfaction will have a positive influence on intent to explore OSDST in current use.	Yes
H12	Satisfaction OSDST will have a positive influence on intent to explore other OSDST.	Yes

*weak support at significance level of 0.1

The results demonstrate that trust in OSDST has very high effect size for satisfaction ($f^2 = 0.799$), indicating greater importance than any other effects in other

significant relationships in the model. Most of the other relationships show medium effect sizes that range from $f^2 = 0.134$ to $f^2 = 0.293$. However, the effect sizes of co-worker championship and satisfaction on their hypothesized dependent variables are small despite their statistical significances at α levels of 0.1 and 0.001 respectively.

Table 21. Effect Sizes (f^2)

	TCOM	SN	TT	Use	Sat	GEXP	CEXP
PC	0.024	0.020					
TCOM		0.134	0.153				
SN			0.293				
TT				0.210	0.799		
Use					0.146		
Sat						0.083	0.040

5.4.2 Post-hoc analysis moderation

One of the motivations of the study is to understand how the users in business units are adapting to working with open source technologies. Therefore, we compare the structural models between the individuals in business and information technology (IT) units in order to discover any underlying differences (Table 22). The PLS multi-group analysis demonstrated nearly identical results in explaining the variance of endogenous variables in the model. In both business and IT units, the explained variance (R^2) was reported as 0.592 and 0.593 respectively. Similarly, 41.7% of the variance in Trust in OSDST is explained for users in business units, while it was 43.5% for the IT workers. However, the multi-group analysis revealed interesting deviations as well:

- Co-Worker Championship → Trust in OSDST Community: For the IT model, the relationship is strongly significant at the significance level of 0.01, while it is not significant for the business model. These findings also provide insight into why this hypothesis was moderately supported for the combined model.
- Satisfaction → Exploration: For the business model, the relationship between OSDST satisfaction and intent to explore the technologies in use is statistically significant at α 0.001. The results reveal the same findings for the influence of OSDST satisfaction on intent to explore other open source technologies. However, the relationship for neither relationship is statistically significant for IT users.
- Trust in OSDST Community → Situational Normality with OSDST: While trust in data science communities is related to the business users' situational normality beliefs strongly ($p < 0.001$), it was not significant for IT users.

Table 22. Group Differences (Business vs IT Units)

	p-Value (Bus)	p-Value (IT)
Co-worker Championship -> Trust in Community	0.914	0.001
Satisfaction -> Explore (Others)	0.000	0.405
Satisfaction -> Explore (Current)	0.000	0.187
Trust in Community -> Situational Normality	0.000	0.105

After the multi-group analysis revealed differences between business and IT groups, we posit the following moderating relationships for the relevant hypotheses:

H4a: The positive influence of co-worker championship on Trust in OSDST community will be moderated by departmental units, such that it will have an influence only for the users in IT units.

H5a: The positive influence of Trust in OSDST community on situational normality will be moderated by departmental units, such that it will have an influence only for users in business units.

H11a: The positive influence of OSDST satisfaction on intent to explore current OSDST will be moderated by departmental units, such that it will have an influence only for users in business units.

H12a: The positive influence of OSDST satisfaction on intent to explore other OSDST will be moderated by departmental units, such that it will have an influence only for users in business units.

Next, we analyzed age and gender as the other categorical control variables. PLS-MGA did not reveal any statistical differences for age groups. However, male and female users differed in the relationship between co-worker championship and trust in OSDST community. For male users, the relationship was significant at the significance level of 0.05 ($p=0.029$). On the other hand, the relationship was not significant for female users. Interestingly, the same relationship was also reported to be moderated by the departmental units, suggesting that it was significant only for IT users. This time, our findings also suggest that it is significant only for female IT users. Accordingly, we revise our moderation hypothesis.

H4a: The positive influence of co-worker championship on Trust in OSDST community will be moderated by departmental units and gender, such that it will influence IT users who are female.

In the next chapter, we summarize the findings of the study, present our contributions for research and practice, identify its limitations, and offer new avenues for future research.

CHAPTER 6

CONCLUSION

6.1 Summary of findings

We have tested our research model with a sample of 173 respondents using the PLS-SEM methodology. The measurement model results indicated that our research instrument had high reliability and validity, ensuring confidence in our hypothesis testing. The structural model results demonstrated that the most of our hypotheses are supported while revealing surprising findings especially during the exploratory post-hoc analysis.

Consistent with the feature-centric view of IT use (Jasperson et al., 2005) and IT continuance model (Bhattacharjee, 2001), we found that satisfaction with OSDST was positively associated with exploration intentions. Satisfied users would not only try to explore more features of the technologies but also put an effort to discover and explore the other OSDST that they have not used yet. However, post-hoc moderation tests reveal an interesting finding: The relationship was dependent on departmental units. While satisfied users at the business units were willing to engage in exploration behavior, we could not see the same outcome for the respondents from IT units. This may be due to their differences in the depth of their technical knowledge, overall familiarity or comfort with technologies in general. Regardless of their satisfaction with the technology they are using, users at IT units may already be willing to explore other technologies. On the other hand, for business users satisfaction would be the strongest indicator of working with a particular technology or other technologies in the same class.

Based on McKnight et al.'s trust in technology models (2009, 2011), we hypothesized that trusting beliefs in OSDST would have a positive influence on perceived usefulness and satisfaction. Both hypotheses are also supported strongly, indicating that trust in technology is a critical factor not only for the initial acceptance but also for the continued use. We have also found strong support for the relationships hypothesized as the antecedents of trust in OSDST. Results confirmed that both situational normality and trust in OSDST community are positively related to trust in OSDST during the post-adoption stage.

On the other hand, we found little support for the hypothesized influence of management and co-worker championship on general trust beliefs of situational normality and trust in OSDST community. This may indicate that championship efforts may lose their influence and no longer be needed once IT is implemented, and the users have completed initial acceptance. However, we should note that co-worker championship was found to be positively associated with trust in OSDST.

6.2 Academic and practical contributions

This research contributes to academy and practice in several ways. To the best of our knowledge, this is the first study to develop and test a theoretical model that investigates the individual use of open source data science technologies in organizations, where most individuals from business units actively adopt and utilize open source technologies for the first time. We have highlighted and examined the users' trusting beliefs in both the data science community and the technology itself and demonstrated their importance in the continued use of OSDST.

Overall, the results of this study sheds light on plausible strategies to be adopted by organizations to encourage the exploration and utilization of OSDST.

This research informs managers on how building trust in OSDST and its supporting community relate to their utilization across the whole organization. Managers should also take into account the impact of building a collaborative environment where the active participation of co-workers may positively influence an individual's trusting beliefs and behaviors.

6.3 Limitations and future research

This study has several limitations that need to be pointed out. First limitation is the cross-sectional nature of the data collection process. Field studies that require the gathering of data from organizations is particularly challenging because of the difficulties involved in reaching the respondents again or respondent dropouts due to employee churn. Despite these difficulties, we believe that our understanding of OSDST utilization would improve greatly if future research tests the model longitudinally.

We gathered our data at a single point of time, implying that we could capture a snapshot of the IT use in organizations. Future research should consider replicating this study with a longitudinal design which would observe the changes and variation in use behavior as well as confirm causality for the hypothesized relationships.

Our sample consisted of OSDST users in a single, developing country. Also, it was somewhat imbalanced in terms of the industries represented. Industries except banking and telecom was underrepresented mainly because large scale data science initiatives are yet to be diffused for them. Our interviews with managers revealed that many firms in these industries still have not adopted open source technologies or even started to use advanced analytics techniques for data utilization. Therefore, the results of this study should carefully be generalized to the whole population of

OSDST users. We suggest researchers to examine the phenomenon in developed countries where data science maturity would be extended to other industries such as production, manufacturing, and energy. We also recommend cross-country or cross-cultural studies that would reveal the differences in organizational practices, perceptions of trust, and exploration behavior. Also, certain demographical information could not be captured in depth due to the sensitivity of some questions in the survey. Future studies that focus on trust building mechanisms but excludes sensitive organizational questions may be helpful.

Although PLS-SEM allows researchers to work with small sample sizes to validate the research instrument and estimate path coefficients, it is suggested that larger sample sizes may improve the precision of bootstrapped algorithms. One of the reasons why some of our relationships are reported to have low statistical significance may be related to our sample size, and future research may use a larger number of respondents to rule this issue out.

Finally, now that we have provided evidence that trust plays a key role in the user satisfaction and exploration intentions, we recommend future studies to incorporate other organizational and social factors that would strengthen the trust beliefs of OSDST users.

APPENDIX A

ETHICS COMMITTEE APPROVAL

T.C.

BOĞAZIÇI ÜNİVERSİTESİ

Sosyal ve Beşerî Bilimler Yüksek Lisans ve Doktora Tezleri Etik İnceleme Komisyonu

Sayı: 2019-36

28 Mart 2019

Rıza Ergün Arsal
İşletme

Sayın Araştırmacı,

"The Acceptance and Continued Usage of Open Source Data Science and Advanced Analytics Technologies" başlıklı projeniz ile ilgili olarak yaptığımız SBB-EAK 2019/4 sayılı başvuru komisyonumuz tarafından 28 Mart 2019 tarihli toplantıda incelenmiş ve uygun bulunmuştur.

Dr. Öğr. Üyesi İnci Aytan

Prof. Dr. Feyza Çorapçı

Doç. Dr. Mehmet Yiğit Özlüdal

Doç. Dr. Ebru Kaya

Dr. Öğr. Üyesi Şebnem Yalçın

APPENDIX B

SURVEY ADMINISTRATION

Sayın Katılımcı,

Açık Kaynak Kodlu Veri Bilimi teknolojilerinin organizasyonlardaki benimsenmesi ve kullanımıyla ilgili gerçekleştirdiğimiz bu çalışmaya sizi davet ediyoruz. Bu araştırma ile bu teknolojilerin benimsenmesinde ve etkin olarak kullanılmasında rol oynayan kritik faktörleri belirlemeyi hedefliyoruz.

Bu araştırma bilimsel bir amaçla yapılmaktadır ve çalışmaya katılımınız tamamen isteğe bağlıdır. Katılımınızın karşılığında bir ödeme ya da ödül verilmemektedir. Ayrıca araştırmaya katılmayı kabul etmiş olsanız bile istediğiniz zaman ve hiçbir yükümlülük olmadan geri çekilme hakkınız bulunmaktadır.

Araştırma süreci boyunca isminiz ya da email adresiniz gibi kişisel bilgileriniz sorulmayacaktır. Ayrıca bütün cevaplarınız kesin gizlilik prensibi ile saklanacak ve korunacaktır. Anket çalışması boyunca toplanan tüm cevaplar anonim olarak değerlendirilecektir ve sadece toplu olarak analiz edilecektir. Sizden alınan bilgiler başka bir çalışma için kullanılmayacak, yöneticileriniz ya da işyerinizden herhangi bir çalışan ile paylaşılmayacaktır.

Herhangi bir sorunuz olması durumunda projenin araştırmacısı Rıza Ergün Arsal (ergun.arsal@bilgi.edu.tr) ve proje yürütücüsü Dr. Öğretim Üyesi Mustafa Hayri Tongarlak (tongarlak@boun.edu.tr) ile iletişime geçebilirsiniz.

Araştırmayla ilgili haklarınız konusunda da Boğaziçi Üniversitesi Etik Kurulu'na (sbinarek@boun.edu.tr) danışabilirsiniz.

Katılımınız çalışmanın başarısı açısından çok önemlidir. Anket sorularını cevaplamanız yaklaşık 15 dakikanızı alacaktır. Değerli zamanınızı ayırdığınız için şimdiden çok teşekkür ederiz.

APPENDIX C

SURVEY INSTRUMENT IN TURKISH

Stem:

Öncelikle sizin bu teknolojileri kullanımınızla ilgili birkaç soru sormak istiyoruz.

- Açık kaynak kodlu veri bilimi teknolojilerini aktif olarak ne kadar zamandır kullanıyorsunuz?.
- Kurumunuzdaki rolünüzde bu teknolojilerinden hangilerini aktif olarak kullanıyorsunuz? Lütfen aşağıdaki boş alana yazınız.

Stem:

Kullanmakta olduğunuz açık kaynak kodlu veri bilimi teknolojileri ile yaşadığınız deneyimlerinizi düşünerek aşağıdaki ifadelere katılma durumunuzu belirtir misiniz?

- Construct: Usefulness:

USE1: Açık kaynak kodlu veri bilimi teknolojilerini kullanmak iş performansımı artırıyor.

USE2: Açık kaynak kodlu veri bilimi teknolojilerini kullanmak veri bilimi ile ilgili işlerimi tamamlamadaki etkinliğimi artırıyor.

USE3: Açık kaynak kodlu veri bilimi teknolojilerini kullanmak veri bilimi projelerini tamamlamadaki verimliliğimi artırıyor.

- Construct: Situational Normality:

SN1. Açık kaynak kodlu veri bilimi teknolojileri kullanma konusunda kendimi son derece rahat hissediyorum.

SN2. Açık kaynak kodlu veri bilimi teknolojileri kullandığım zaman kendimi çok iyi hissediyorum.

SN3. Açık kaynak kodlu veri bilimi teknolojileri kullandığım zaman işlerim yolunda gidecek gibi hissediyorum.

SN4. Açık kaynak kodlu veri bilimi teknolojileri kullandığım zaman beklediğim sonuçları alacağıma dair güvenim tam.

- Construct: Trust in OSDST - Functionality

TTF1: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojileri, ihtiyacım olan işlevselliğe sahiptir.

TTF2: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojileri, veri bilimi ile ilgili işlerimi gerçekleştirmek için gerekli olan özelliklere sahiptir.

TTF3: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojileri, işimde ihtiyacım olan kabiliyetlere sahiptir.

- Construct: Trust in OSDST - Helpfulness

TTH1: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin yardım işlevleri, bana yeterli yönlendirmeyi sağlıyor.

TTH2: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin yardım işlevleri bana ihtiyacım olan her türlü yardımı sağlıyor.

TTH3: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin yardım işlevleri, gerektiğinde çok etkili tavsiyelerde bulunuyor.

TTH4: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojileri, yardım ihtiyacımı çağrılabilir bir yardım fonksiyonu aracılığıyla sağlıyor.

- Construct: Trust in OSDST - Predictability

TTP1: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojileri öngörülebilir bir şekilde davranıyor.

TTP2: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin nasıl çalışacağını önceden tahmin edebilirim.

TTP3: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin işlevleri her kullandığımda aynı şekilde çalışmıyor.

TTP4: Bir çalışma aracı olarak, şu anda kullandığım açık kaynak kodlu veri bilimi teknolojilerinin son derece öngörülebilir olduğunu düşünüyorum.

- Construct: Satisfaction

SAT1: Şu ana dek kullandığım açık kaynak kodlu veri bilimi teknolojileri beni son derece tatmin etti.

SAT2: Şu ana dek kullandığım açık kaynak kodlu veri bilimi teknolojilerinden son derece memnun kaldım.

SAT3: Şu ana dek kullandığım açık kaynak kodlu veri bilimi teknolojileri ihtiyaçlarımı fazlasıyla karşıladı.

SAT4: Şu ana dek kullandığım açık kaynak kodlu veri bilimi teknolojileri beklentilerimi fazlasıyla aştı.

- Construct: Intent to Explore (General)

GEXP1: Veri bilimi ile ilgili farklı yöntemleri öğrenmek için çeşitli açık kaynak kodlu veri bilimi teknolojilerini kullanmayı deniyorum.

GEXP2: Veri bilimi konusundaki becerilerimi geliştirebilmek için farklı açık kaynak kodlu veri bilimi teknolojilerini araştırıyorum.

GEXP3: Veri bilimi için kullanılacak açık kaynak kodlu yeni veri bilimi teknolojilerini keşfetmek için hatırı sayılır bir zaman harcıyorum.

GEXP4: Açık kaynak kodlu farklı veri bilimi teknolojilerini keşfetmek için ciddi bir çaba sarfediyorum.

- Construct: Intent to Utilize (Current)

CEXP1: Yararlanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerinin farklı kullanım biçimlerini araştırmak için ciddi bir zaman harcama niyetindeyim.

CEXP2: İşteki verimliliğimi artırmak için, kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerini derinlemesine keşfetmeyi planlıyorum.

CEXP3: Kullanmakta olduğum açık kaynak kodlu veri bilimi teknolojilerini yeni biçimlerde de kullanmayı düşünüyorum.

Stem:

Açık kaynak kodlu veri bilimi teknolojileri, ticari yazılım desteğine sahip olmamakla beraber bu teknolojileri kullanan insanların oluşturduğu komünite (community) desteği ile bu açığı kapatmaktadır. Bu komünitelere örnek vermek gerekirse: Python/R forumları, Kaggle, Stack Overflow, Data Science Central gibi online kaynakları sayabiliriz. Şimdi size bu komünitelerle ilgili fikirlerinizi soracağız.

- Trust in OSDST Community

TCOM1: Bu komünitelerde açık kaynak kodlu veri bilimi teknolojilerinin kullanımı ile ilgili yardımcı olabilecek yeterli kaynağın olduğunu düşünüyorum.

TCOM2: İhtiyacım olduğunda açık kaynak kodlu veri bilimi teknolojileri ile ilgili güvenilir kaynakları bu komünitelerde bulabilirim.

TCOM3: Açık kaynak kodlu veri bilimi teknolojileriyle ilgili ihtiyacım olan bilgileri bu komünitelerde bulabilme konusunda kendimi rahat hissediyor. İhtiyacım olsa bu komünitelerden birilerinin bana yardımcı olacağına inanıyorum.

TCOM4: İhtiyacım olsa bu komünitelerden birilerinin bana yardımcı olacağına inanıyorum.

Stem:

Açık kaynak kodlu veri bilimi teknolojilerinin kullanımında kurumsal faktörlerin rolünü anlamak istiyoruz. O yüzden bu teknolojilerle ilgili üst yönetiminizin yaklaşımını bilmek bizim için çok önemli. Aşağıdaki ifadelerle katılma durumunuzu kurumunuzun üst yönetimini düşünerek bizimle paylaşır mısınız?

- Top Management Championship - Belief

TCB1: Üst yönetimimiz açık kaynak kodlu veri bilimi teknolojilerinin kurumumuz için önemli bir iş değeri yaratma potansiyeline sahip olduğuna inanıyor.

TCB2: Üst yönetimimiz açık kaynak kodlu veri bilimi teknolojilerinin şirketler için yeni bir rekabet alanı yaratacağını düşünüyor.

TCB3: Üst yönetimimiz iş faaliyetlerimizi sürdürebilmek için açık kaynak kodlu veri bilimi teknolojilerinin gerekli olmadığına inanıyor.

- Top Management Championship - Participation

TCP1: Üst yönetimimiz açık kaynak kodlu veri bilimi teknolojilerinin kurumumuzdaki kullanımı için bir vizyon geliştirmiştir.

TCP2: Üst yönetimimiz açık kaynak kodlu veri bilimi teknolojilerinin kurumumuzdaki kullanımı için bir strateji geliştirmiştir.

TCP3: Üst yönetimimiz açık kaynak kodlu veri bilimi teknolojileri ile geliştirilen projelerin değerlendirilmesi için gerekli hedef ve standartları belirlemiştir.

Stem:

Aşağıdaki soruları bu kez de kurumunuzdaki çalışma arkadaşlarınızı düşünerek yanıtlar mısınız?

- Co-worker Championship - Belief

PCB1: Çalışma arkadaşlarım açık kaynak kodlu veri bilimi teknolojilerinin şirketimiz için önemli bir iş değeri yaratma potansiyeline sahip olduğuna inanıyorlar.

PCB2: Çalışma arkadaşlarım açık kaynak kodlu veri bilimi teknolojilerinin şirketler için yeni bir rekabet alanı yaratacağına inanmıyorlar.

PCB3: Çalışma arkadaşlarım iş faaliyetlerimizi sürdürebilmek için açık kaynak kodlu veri bilimi teknolojilerinin gerekli olduğunu düşünüyorlar.

- Co-worker Championship - Participation

PCP1: Çalışma arkadaşlarım veri bilimi ve analitiği konularındaki vasıflarını geliştirmek için farklı açık kaynak kodlu veri bilimi teknolojilerini öğrenirler.

PCP2: Çalışma arkadaşlarım yeni projelerin geliştirme ve/veya test süreçlerinde açık kaynak kodlu veri bilimi teknolojilerini de denerler.

PCP3: Çalışma arkadaşlarım canlı ortamdaki uygulamalar için açık kaynak kodlu veri bilimi teknolojilerini de kullanırlar.

Stem:

Artık araştırmamızın sonuna geliyoruz. Birkaç soru ile sizi biraz daha yakından tanıyalım istiyoruz.

Kurumunuzdaki Yeriniz (İş Birimi/Bilgi Teknolojileri (BT))

Cinsiyetiniz (Erkek/Kadın)

Yaş Aralığımız (18-24/25-34/35-44/45-54/55-64/65-74/75 ve üzeri)

Eğitim Seviyeniz (Lise veya altı/Üniversite (Lisans)/Yüksek Lisans (Master)/Doktora)

Çalıştığınız Endüstri

APPENDIX D

NORMAL DISTRIBUTION MEASURES

Item	Min	Max	Mean	Median	Standard Deviation	Skewness	Kurtosis
MCB1	1	7	5.715	6	1.236	-1.211	1.487
MCB2	1	7	5.671	6	1.216	-1.013	1.173
MCB3	1	7	5.576	6	1.640	-1.338	1.042
MCP1	1	7	5.250	6	1.411	-0.845	0.232
MCP2	1	7	5.069	5	1.535	-0.703	-0.256
MCP3	1	7	4.740	5	1.565	-0.508	-0.403
PCB1	1	7	5.815	6	1.162	-1.412	2.885
PCB2	1	7	5.017	6	1.942	-0.747	-0.764
PCB3	1	7	5.705	6	1.257	-1.112	1.225
PCP1	1	7	5.355	6	1.489	-1.154	1.148
PCP2	1	7	5.272	6	1.463	-0.955	0.669
PCP3	1	7	5.127	6	1.669	-0.803	-0.147
TSN1	1	7	5.424	6	1.243	-0.889	1.128
TSN2	3	7	5.965	6	1.017	-0.837	0.192
TSN3	2	7	5.618	6	1.070	-0.544	-0.047
TSN4	2	7	5.384	5	1.039	-0.286	-0.091
TCOM1	1	7	5.792	6	1.058	-1.035	2.197
TCOM2	2	7	5.908	6	0.935	-0.762	0.980
TCOM3	2	7	5.855	6	0.998	-0.699	0.519
TCOM4	2	7	5.931	6	1.003	-0.873	0.768
TTF1	2	7	5.831	6	0.962	-0.853	1.208
TTF2	4	7	5.983	6	0.862	-0.521	-0.385
TTF3	3	7	5.988	6	0.821	-0.743	0.663
TTH1	2	7	5.370	5	1.041	-0.540	0.471
TTH2	2	7	5.105	5	1.224	-0.396	-0.190

TTH3	2	7	4.895	5	1.229	-0.219	-0.506
TTH4	1	7	4.960	5	1.348	-0.473	-0.099
TTP1	2	7	5.543	6	0.955	-0.712	0.942
TTP2	3	7	5.186	5	1.065	-0.291	-0.346
TTP3	1	7	4.983	5	1.305	-0.476	-0.231
TTP4	2	7	5.387	5	0.980	-0.543	0.746
PU1	3	7	6.238	6	0.896	-1.033	0.476
PU2	4	7	6.289	6	0.805	-0.979	0.401
PU3	1	7	6.318	7	0.881	-1.960	7.119
SAT1	3	7	5.895	6	0.831	-0.667	0.769
SAT2	3	7	5.848	6	0.854	-0.619	0.509
SAT3	3	7	5.676	6	0.940	-0.543	0.170
SAT4	2	7	5.231	5	1.091	-0.282	-0.498
GEXP1	1	7	5.506	6	1.420	-1.096	0.819
GEXP2	1	7	5.514	6	1.379	-1.026	0.665
GEXP3	1	7	4.794	5	1.595	-0.375	-0.701
GEXP4	1	7	4.746	5	1.590	-0.532	-0.597
CEXP1	1	7	5.214	5	1.425	-0.774	0.049
CEXP2	1	7	5.637	6	1.240	-1.190	1.585
CEXP3	2	7	5.570	6	1.185	-0.712	0.197

APPENDIX E
OUTER LOADINGS

	MCB	MCP	PCB	PCP	TSN	TCOM	TTP	TTF	TTH	PU	SAT	GEXP	CEXP
MCB1	0.942												
MCB2	0.944												
MCP1		0.926											
MCP2		0.950											
MCP3		0.898											
PCB1			0.930										
PCB3			0.931										
PCP1				0.913									
PCP2				0.949									
PCP3				0.883									
TSN1					0.809								
TSN2					0.795								
TSN3					0.850								
TSN4					0.755								
TCOM1						0.858							
TCOM2						0.916							
TCOM3						0.915							
TCOM4						0.826							
TTP1							0.846						
TTP2							0.839						

TTP4							0.883						
TTF1								0.871					
TTF2								0.924					
TTF3								0.869					
TTH1									0.897				
TTH2									0.895				
TTH3									0.839				
TTH4									0.625				
PU1										0.885			
PU2										0.904			
PU3										0.910			
SAT1											0.906		
SAT2											0.912		
SAT3											0.894		
SAT4											0.782		
GEXP1												0.888	
GEXP2												0.887	
GEXP3												0.855	
GEXP4												0.796	
CEXP1													0.856
CEXP2													0.929
CEXP3													0.872

APPENDIX F

DISCRIMINANT VALIDITY: CROSS LOADINGS

	MCB	MCP	PCB	PCP	TSN	TCOM	TTP	TTF	TTH	PU	SAT	GEXP	CEXP
MCB1	0.942	0.715	0.576	0.478	0.103	0.122	0.031	0.066	0.149	0.065	0.040	0.029	0.115
MCB2	0.944	0.733	0.443	0.357	0.116	0.157	0.104	0.097	0.210	0.086	0.118	0.052	0.116
MCP1	0.736	0.926	0.446	0.416	0.166	0.101	0.124	0.119	0.187	0.100	0.135	0.109	0.155
MCP2	0.737	0.950	0.429	0.398	0.162	0.141	0.150	0.077	0.224	0.039	0.117	0.074	0.173
MCP3	0.654	0.898	0.443	0.418	0.030	0.105	0.033	-0.006	0.230	-0.010	0.025	0.137	0.113
PCB1	0.557	0.486	0.930	0.733	0.193	0.249	0.192	0.179	0.208	0.280	0.186	0.115	0.160
PCB3	0.449	0.399	0.931	0.740	0.243	0.248	0.204	0.197	0.213	0.192	0.146	0.051	0.080
PCP1	0.401	0.408	0.744	0.913	0.194	0.088	0.112	0.139	0.180	0.212	0.123	0.070	0.141
PCP2	0.369	0.393	0.727	0.949	0.209	0.167	0.189	0.153	0.230	0.187	0.132	0.080	0.045
PCP3	0.449	0.417	0.703	0.883	0.138	0.138	0.151	0.079	0.161	0.138	0.108	0.053	-0.013
TSN1	0.103	0.101	0.227	0.211	0.809	0.424	0.527	0.467	0.278	0.498	0.533	0.364	0.300
TSN2	0.026	0.113	0.131	0.074	0.795	0.218	0.326	0.376	0.131	0.448	0.386	0.358	0.372
TSN3	0.094	0.094	0.161	0.124	0.850	0.289	0.408	0.412	0.221	0.430	0.445	0.278	0.371
TSN4	0.134	0.118	0.211	0.194	0.755	0.215	0.386	0.378	0.288	0.341	0.363	0.129	0.252
TCOM1	0.078	0.058	0.225	0.118	0.331	0.858	0.397	0.335	0.280	0.207	0.319	0.107	0.068
TCOM2	0.180	0.155	0.265	0.143	0.333	0.916	0.402	0.339	0.353	0.197	0.370	0.084	0.127
TCOM3	0.136	0.121	0.206	0.103	0.365	0.915	0.392	0.319	0.336	0.241	0.420	0.148	0.181
TCOM4	0.125	0.102	0.244	0.142	0.285	0.826	0.406	0.269	0.291	0.206	0.297	0.076	0.185
TTP1	0.061	0.105	0.133	0.132	0.394	0.366	0.846	0.474	0.411	0.364	0.535	0.182	0.204
TTP2	0.058	0.067	0.183	0.152	0.438	0.445	0.839	0.342	0.271	0.278	0.404	0.151	0.140

TTP4	0.065	0.112	0.234	0.141	0.527	0.363	0.883	0.438	0.331	0.337	0.458	0.202	0.271
TTF1	0.081	0.075	0.160	0.135	0.528	0.268	0.415	0.871	0.414	0.411	0.652	0.210	0.228
TTF2	0.095	0.054	0.191	0.124	0.405	0.327	0.468	0.924	0.343	0.340	0.653	0.223	0.338
TTF3	0.054	0.057	0.187	0.103	0.443	0.364	0.431	0.869	0.374	0.415	0.695	0.262	0.344
TTH1	0.128	0.157	0.222	0.180	0.268	0.338	0.332	0.451	0.897	0.155	0.402	0.163	0.082
TTH2	0.170	0.176	0.203	0.255	0.305	0.245	0.400	0.426	0.895	0.211	0.437	0.212	0.096
TTH3	0.218	0.281	0.189	0.164	0.204	0.317	0.305	0.298	0.839	0.114	0.321	0.171	0.027
TTH4	0.110	0.155	0.110	0.041	0.174	0.306	0.260	0.149	0.625	-0.053	0.160	0.093	0.085
PU1	0.128	0.063	0.264	0.205	0.461	0.208	0.350	0.385	0.146	0.885	0.433	0.245	0.206
PU2	0.046	0.010	0.176	0.155	0.474	0.198	0.339	0.374	0.109	0.904	0.454	0.201	0.216
PU3	0.046	0.054	0.242	0.171	0.519	0.242	0.348	0.418	0.149	0.910	0.526	0.227	0.166
SAT1	0.058	0.069	0.167	0.060	0.541	0.329	0.548	0.683	0.328	0.480	0.906	0.244	0.316
SAT2	0.034	0.041	0.123	0.098	0.496	0.346	0.469	0.676	0.360	0.499	0.912	0.237	0.296
SAT3	0.130	0.168	0.185	0.174	0.497	0.401	0.474	0.723	0.432	0.501	0.894	0.227	0.315
SAT4	0.071	0.073	0.148	0.133	0.382	0.328	0.422	0.526	0.347	0.346	0.782	0.270	0.258
GEXP1	-0.014	0.055	0.066	0.066	0.368	0.125	0.268	0.240	0.189	0.278	0.291	0.888	0.461
GEXP2	0.059	0.118	0.127	0.123	0.338	0.114	0.208	0.264	0.225	0.242	0.238	0.887	0.472
GEXP3	0.063	0.132	0.067	0.022	0.259	0.113	0.106	0.226	0.129	0.157	0.235	0.855	0.616
GEXP4	0.062	0.099	0.033	0.033	0.216	0.024	0.088	0.134	0.126	0.146	0.149	0.796	0.600
CEXP1	0.095	0.107	0.094	-0.006	0.302	0.128	0.182	0.237	0.042	0.167	0.251	0.651	0.856
CEXP2	0.161	0.224	0.133	0.094	0.387	0.168	0.247	0.355	0.091	0.195	0.339	0.521	0.929
CEXP3	0.062	0.079	0.110	0.065	0.360	0.121	0.207	0.300	0.093	0.211	0.303	0.476	0.872

APPENDIX G

MULTICOLLINEARITY – OUTER VIF VALUES

Items	VIF
CEXP1	2.237
CEXP2	2.857
CEXP3	2.040
GEXP1	3.048
GEXP2	3.222
GEXP3	3.655
GEXP4	3.219
MCB1	2.532
MCB1	2.916
MCB2	2.532
MCB2	3.020
MCP1	3.570
MCP1	3.916
MCP2	4.544
MCP2	4.784
MCP3	2.691
MCP3	2.763
PCB1	2.161
PCB1	2.650
PCB3	2.161
PCB3	2.694
PCP1	3.311

PCP1	3.802
PCP2	4.465
PCP2	4.517
PCP3	2.445
PCP3	2.744
PU1	2.320
PU2	2.646
PU3	2.502
SAT1	4.211
SAT2	4.234
SAT3	2.738
SAT4	1.836
TCOM1	2.381
TCOM2	3.756
TCOM3	3.669
TCOM4	2.078
TSN1	1.521
TSN2	1.930
TSN3	2.234
TSN4	1.595
TTF1	2.165
TTF1	2.351
TTF2	2.999
TTF2	3.285
TTF3	2.164
TTF3	2.444
TTH1	2.833

TTH1	3.092
TTH2	2.948
TTH2	3.130
TTH3	2.027
TTH3	2.076
TTH4	1.329
TTH4	1.443
TTP1	1.652
TTP1	1.970
TTP2	1.881
TTP2	1.986
TTP4	2.098
TTP4	2.203

REFERENCES

- Agarwal, R., Sambamurthy, V., & Stair, R. M. (2000). The evolving relationship between general and specific computer self-efficacy—An empirical assessment. *Information Systems Research*, 11(4), 418-430.
- Albers, S. (2010). PLS and success factor studies in marketing. In *Handbook of Partial Least Squares* (pp. 409-425). Springer, Berlin, Heidelberg.
- Alreck, P. L., & Settle, R. B. (1995). *The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey*. New York, NY: IRWIN Professional Publishing.
- Augur, H. (2016). *The Future of Big Data is Open Source*. Retrieved from [http://dataconomy.com/the-future-of-big-data-is-open source/](http://dataconomy.com/the-future-of-big-data-is-open-source/)
- Awan A.A., (2022). *Top Data Science Tools for 2022*. Retrieved from <https://www.kdnuggets.com/2022/03/top-data-science-tools-2022.html>
- Bean, R. (2014). Why becoming a data-driven organization is so hard. *Harvard Business Review*, Retrieved from <https://hbr.org/2022/02/why-becoming-a-data-driven-organization-is-so-hard>
- Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3), 4.
- Beyer, M. (2011). *Gartner Says Solving “Big Data” Challenge Involves More Than Just Managing Volumes of Data*. Retrieved from <https://www.gartner.com/newsroom/id/1731916>
- Bhattacharjee, A. (2001). An empirical analysis of the antecedents of electronic commerce service continuance. *Decision Support Systems*, 32(2), 201-214.
- Bhattacharjee, A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly*, 25(3), 351-370.
- Bhattacharjee, A. (2012). *Social science research: Principles, methods, and practices*. University of South Florida.

- Booz Allen Hamilton (2017). *Data science playbook*. Retrieved from:
<https://www.boozallen.com/s/insight/publication/data-science-playbook.html>
- Borgman, H. P., Bahli, B., Heier, H., & Schewski, F. (2013). Cloudrise: exploring cloud computing adoption and governance with the TOE framework. In *46th Hawaii International Conference on System Sciences* (pp. 4425-4435).
- Bruner II, G. C., & Kumar, A. (2005). Explaining consumer acceptance of handheld Internet devices. *Journal of Business Research*, *58*(5), 553-558.
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decision making affect firm performance? Available at SSRN 1819486.
- Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, *26*(4), 435-446.
- Chatterjee, D., Grewal, R., & Sambamurthy, V. (2002). Shaping up for e-commerce: institutional enablers of the organizational assimilation of web technologies. *MIS Quarterly*, *26*(2), 65-89.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, *295*(2), 295-336.
- Chin, W. W., & Newsted, P. R. (1995). The importance of specification in causal modeling: The case of end-user computing satisfaction. *Information Systems Research*, *6*(1), 73-81.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.
- Cooper, R. B., & Zmud, R. W. (1990). Information technology implementation research: a technological diffusion approach. *Management Science*, *36*(2), 123-139.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Davis Jr, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology).

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35(8), 982-1003.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22(14), 1111-1132.
- Dedrick, J., & West, J. (2003, December). Why firms adopt open source platforms: A grounded theory of innovation and standards adoption. In *Proceedings of the workshop on standard making: A critical research frontier for information systems* (pp. 236-257). Seattle, WA.
- Del Bianco, V., Lavazza, L., Morasca, S., & Taibi, D. (2011). A survey on open source software trustworthiness. *IEEE Software*, 28(5), 67-75.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems*, 19(4), 9-30.
- DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5(2), 121-147.
- Edison (2017). Edison Data Science Framework. *EDISON Project*, Retrieved from <https://edison-project.eu/edison/edison-data-science-framework-edsf/>
- Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research*, 28(4), 854-887
- Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, Mass: Addison-Wesley Pub.
- Fitzgerald, B. (2006). The transformation of open source software. *MIS Quarterly*, 30(3), 587-598.
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3), 382-388

- Francisco, K., & Swanson, D. (2018). The supply chain has no clothes: Technology adoption of blockchain for supply chain transparency. *Logistics*, 2(1), 2.
- Gallego, M. D., Bueno, S., Racero, F. J., & Noyes, J. (2015). Open source software: The effects of training on acceptance. *Computers in Human Behavior*, 49, 390-399.
- Gallego, M. D., Luna, P., & Bueno, S. (2008). User acceptance model of open source software. *Computers in Human Behavior*, 24(5), 2199-2216.
- Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information systems*, 16(1), 5.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51-90.
- Gefen, D., Straub, D., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4(1), 7.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 25(2), 186-192.
- Giddens, A. (1984). *The constitution of society: outline of the theory of structuration*. University of California Press, Berkeley, CA
- Gopalakrishna-Remani, V., Jones, R. P., & Camp, K. M. (2019). Levels of EMR adoption in US hospitals: An empirical examination of absorptive capacity, institutional pressures, top management beliefs, and participation. *Information Systems Frontiers*, 21(6), 1325-1344.
- Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388-423.
- Gwebu, K. L., & Wang, J. (2011). Adoption of Open Source Software: The role of social identification. *Decision Support Systems*, 51(1), 220-229.
- Hair, J., et al. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd ed., SAGE Publications.

- Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems*, 117(3), 442-458.
- Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010) *Multivariate Data Analysis*. 7th Edition, Pearson, New York.
- Hameed, M. A., Counsell, S., & Swift, S. (2012). A conceptual model for the process of IT innovation adoption in organizations. *Journal of Engineering and Technology Management*, 29(3), 358-390.
- Harris, D. (2013). *The History of Hadoop: From 4 Nodes to the Future of Data*. Retrieved from: <https://old.gigaom.com/2013/03/04/the-history-of-hadoop-from-4-nodes-to-the-future-of-data/>
- Hauge, Ø., Ayala, C., & Conradi, R. (2010). Adoption of open source software in software-intensive organizations—A systematic literature review. *Information and Software Technology*, 52(11), 1133-1154.
- Holden, R. J., & Karsh, B. T. (2010). The technology acceptance model: its past and its future in health care. *Journal of Biomedical Informatics*, 43(1), 159-172.
- Hsieh, J. J., & Wang, W. (2007). Explaining employees' extended use of complex information systems. *European Journal of Information Systems*, 16(3), 216-227.
- Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information Technology and Management*, 1(1), 45-71.
- Jasperson, J. S., Carter, P. E., & Zmud, R. W. (2005). A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. *MIS Quarterly*, 29(3), 525-557.
- Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information Technology*, 21(1), 1-23.
- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly*, 23(2), 183-213.

- Kijsanayotin, B., Pannarunothai, S., & Speedie, S. M. (2009). Factors influencing health information technology adoption in Thailand's community health centers: Applying the UTAUT model. *International Journal of Medical Informatics*, 78(6), 404-416.
- Kim, S. S., & Malhotra, N. K. (2005). A longitudinal model of continued IS use: An integrative view of four mechanisms underlying postadoption phenomena. *Management Science*, 51(5), 741-755.
- Kline, R. B. (1998). *Structural equation modeling*. New York: Guilford.
- Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(1), 1-13.
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1-10.
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS- SEM: The inverse square root and gamma- exponential methods. *Information Systems Journal*, 28(1), 227-261.
- Kuan, K. K., & Chau, P. Y. (2001). A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Information & Management*, 38(8), 507-521.
- Lankton, N., McKnight, D. H., & Thatcher, J. B. (2014). Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems*, 23(2), 128-145.
- Lauterbach, J., & Mueller, B. (2014, December). Adopt, adapt, enact or use?. In *Working Conference on Information Systems and Organizations* (pp. 8-29). Springer, Berlin, Heidelberg.
- Lee, K. C., Kang, I., & McKnight, D. H. (2007). Transfer from offline trust to key online perceptions: an empirical study. *IEEE Transactions on Engineering Management*, 54(4), 729-741.
- Li, Y., Tan, C. H., Xu, H., & Teo, H. H. (2011). Open source software adoption: motivations of adopters and amotivations of non-adopters. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 42(2), 76-94.

- Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, 31(1), 59-87.
- Limayem, M., Hirt, S. G., & Cheung, C. M. (2007). How habit limits the predictive power of intention: The case of information systems continuance. *MIS Quarterly*, 31(4), 705-737.
- Lippert, S. K., & Forman, H. (2005). Utilization of information technology: Examining cognitive and experiential factors of post-adoption behavior. *IEEE Transactions on Engineering Management*, 52(3), 363-381.
- Malhotra, M. K., & Grover, V. (1998). An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, 16(4), 407-425.
- Malone, K., & Wolski, R. (2020). Doing data science on the shoulders of giants: the value of open source software for the data science community. *Harvard Data Science Review* <https://hdsr.mitpress.mit.edu/pub/xsrt4zs2/release/4> (31 May 2020).
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H.. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- Maruping, L., Magni, M., Caporarello, L., & Basaglia, S. (2008). What's the weather like? The effect of team climate and individual attributes on individual intention to explore a new technology. *AMCIS 2008 Proceedings*. 383.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- McKnight, D. H. (2005). Trust in information technology. *The Blackwell encyclopedia of management*, 7, 329-331.
- McKnight, D. H., & Chervany, N. L. (2001). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, 6(2), 35-59.

- McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2), 1-25.
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334-359.
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The Journal of Strategic Information Systems*, 11(3-4), 297-323.
- McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, 23(3), 473-490.
- McKnight, H., Carter, M., & Clay, P. (2009). Trust in technology: Development of a set of constructs and measures. *DIGIT 2009 Proceedings*. 10.
- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192-222.
- Murallie, T. (2021). *Welcome to the age of citizen data scientists*. Retrieved from <https://towardsdatascience.com/how-to-become-a-citizen-data-scientist-294660da0494>
- Nagy, D., Yassin, A. M., & Bhattacharjee, A. (2010). Organizational adoption of open source software: barriers and remedies. *Communications of the ACM*, 53(3), 148-151.
- Nambisan, S., Agarwal, R., & Tanniru, M. (1999). Organizational mechanisms for enhancing user innovation in information technology. *MIS Quarterly*, 23(3), 365-395.
- Nelson, R. R. (1991). Educational needs as perceived by IS and end-user personnel: A survey of knowledge and skill requirements. *MIS Quarterly*, 15(4), 503-525.
- Neuman, W. L., & Robson, K. (2014). *Basics of social research*. Toronto: Pearson Canada.

- Noyes, K. (2014). *How a little open source project came to dominate big data*. Retrieved from <https://fortune.com/2014/06/30/hadoop-how-open-source-project-dominate-big-data/>
- Nunnally, J. C. (1967). *Psychometric theory*. McGraw-Hill.
- Nunnally, J.C. and Bernstein, I.H. (1994) The Assessment of Reliability. *Psychometric Theory*, 3, 248-292.
- Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation*, 14(1), 110-121.
- Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management*, 51(5), 497-510.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460-469.
- Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11(4), 404-428.
- Orlikowski, W. J., Yates, J., Okamura, K., & Fujimoto, M. (1995). Shaping electronic communication: The metastructuring of technology in the context of use. *Organization Science*, 6(4), 423-444.
- Park, S. Y. (2009). An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Journal of Educational Technology & Society*, 12(3), 150-162.
- Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7(3), 101-134.
- Pennington, R., Wilcox, H. D., & Grover, V. (2003). The role of system trust in business-to-consumer transactions. *Journal of Management Information Systems*, 20(3), 197-226.

- Piatetsky, G. (2019). *Python leads the 11 top data science, machine learning platforms: Trends and analysis*. Retrieved from <https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>
- Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., & Pahnla, S. (2004). Consumer acceptance of online banking: an extension of the technology acceptance model. *Internet Research*.
- Poole, M. S., & DeSanctis, G. (2004). Structuration theory in information systems research: Methods and controversies. In *The Handbook of Information Systems Research* (pp. 206-249). IGI Global.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109-130.
- Redman, T.C. & Davenport, T.H. (2021). 4 ways to democratize data science in your organization. *Harvard Business Review*, Retrieved from <https://hbr.org/2021/03/4-ways-to-democratize-data-science-in-your-organization>
- Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332-344.
- Rezvani, A., Khosravi, P., & Dong, L. (2017). Motivating users toward continued usage of information systems: Self-determination theory perspective. *Computers in Human Behavior*, 76, 263-275.
- Rogers, E. M. (1995). *Diffusion of innovations*. New York: Free Press.
- Rossi, B., Russo, B., & Succi, G. (2012). Adoption of free/libre open source software in public organizations: factors of impact. *Information Technology & People*.
- Saga, V. L. & Zmud, R. W. (1994), 'The nature and determinants of IT acceptance, routinization and infusion', In *Diffusion, transfer and implementation of information technology* (Ed, Levine, L.), Elsevier Science BV, Amsterdam, pp. 67-86.
- Salancik, G. R., & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23(2), 224-253.

- San Martín, H., & Herrero, Á. (2012). Influence of the user's psychological factors on the online purchase intention in rural tourism: Integrating innovativeness to the UTAUT framework. *Tourism Management*, 33(2), 341-350.
- Schryen, G. (2013). Revisiting IS business value research: what we already know, what we still need to know, and how we can get there. *European Journal of Information Systems*, 22(2), 139-169.
- Sharma, S., Singh, G., Jones, P., Kraus, S., & Dwivedi, Y. K. (2022). Understanding agile innovation management adoption for SMEs. *IEEE Transactions on Engineering Management*. doi: 10.1109/TEM.2022.3148341.
- Silic, M., & Back, A. (2016). The influence of risk factors in decision-making process for open source software adoption. *International Journal of Information Technology & Decision Making*, 15(01), 151-185.
- Soh, C., & Markus, M. L. (1995). How IT creates business value: a process theory synthesis. *ICIS 1995 Proceedings*, 4.
- Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of Marketing*, 60(3), 15-32.
- Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, 13(2), 147-169.
- Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information systems*, 13(1), 24.
- Taylor, S., & Todd, P. (1995). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing*, 12(2), 137-155.
- Thatcher, J. B., McKnight, D., Baker, E. W., Arsal, R. E., & Roberts, N. H. (2011). The role of trust in postadoption it exploration: An empirical examination of knowledge management systems. *IEEE Transactions on Engineering Management*, 58(1), 56-70.
- Thong, J. Y., Hong, S. J., & Tam, K. Y. (2006). The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of Human-Computer Studies*, 64(9), 799-810.

- Tornatzky, L. G., & Fleischer, M. (1990). *Processes of technological innovation*. Lexington Books.
- Van Raaij, E. M., & Schepers, J. J. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education, 50*(3), 838-852.
- Vance, A., Elie-Dit-Cosaque, C., & Straub, D. W. (2008). Examining trust in information technology artifacts: the effects of system quality and culture. *Journal of Management Information Systems, 24*(4), 73-100.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.
- Venkatesh, V., Thong, J. Y., Chan, F. K., Hu, P. J. H., & Brown, S. A. (2011). Extending the two-stage information systems continuance model: Incorporating UTAUT predictors and the role of context. *Information Systems Journal, 21*(6), 527-555.
- Wade, M., & Hulland, J. (2004). The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly, 28*(1), 107-142.
- Wang, W., & Benbasat, I. (2008). Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. *Journal of Management Information Systems, 24*(4), 249-273.
- Wingreen, S. C., & Baglione, S. L. (2005). Untangling the Antecedents and Covariates of E-Commerce Trust: Institutional Trust vs. Knowledge-Based Trust. *Electronic Markets, 15*(3), 246-260.
- Wingreen, S. C., Mazey, N. C., Baglione, S. L., & Storholm, G. R. (2019). Transfer of electronic commerce trust between physical and virtual environments: experimental effects of structural assurance and situational normality. *Electronic Commerce Research, 19*(2), 339-371.
- Woodie, A. (2016). *An Open Source Tour de Force at Apache: Big Data 2016*. Retrieved from <http://www.datanami.com/2016/05/11/open-source-tour-de-force-apache-big-data-2016/>

- Zhou, T., Lu, Y., & Wang, B. (2010). Integrating TTF and UTAUT to explain mobile banking user adoption. *Computers in Human Behavior*, 26(4), 760-767.
- Zhu, K., Kraemer, K. L., & Dedrick, J. (2004). Information technology payoff in e-business environments: An international perspective on value creation of e-business in the financial services industry. *Journal of Management Information Systems*, 21(1), 17-54.
- Zillner, S., Rusitschka, S., & Skubacz, M. (2014, March). Big data story: Demystifying big data with special focus on and examples from industrial sectors. In *EU BIG European Big Data Public Private Forum, White paper* (Vol. 3).
- Zimmer, J. C., Arsal, R., Al-Marzouq, M., Moore, D., & Grover, V. (2010). Knowing your customers: Using a reciprocal relationship to enhance voluntary information disclosure. *Decision Support Systems*, 48(2), 395-406.