Research Proposal for the PhD Scholarship in Global Studies, Economy, Society and Law
Università degli Studi di Urbino Carlo Bo

Title: DATA-DRIVEN INNOVATION: LEVERAGING DIGITAL TRANSFORMATION FOR COMPETITIVENESS IN ICT-USER SECTORS IN EUROPE

Keywords: data-driven innovation, ICT- user industries, regression trees, machine learning
Research area: SECS-P/06 Applied economics

Candidate Full Name: Luis Carlos Castillo-Téllez
Latest Academic Degree: Master in Economic Theory, Istanbul University

GENERAL PRESENTATION OF THE PROJECT AND STATE OF THE ART

The diffusion and adoption of ubiquitous digital technologies that allow the generation, accumulation, processing and analysis of vast amounts of data has become a fundamental source of competitive advantage and value creation for firms (Barham, 2017). In the current context of rapid technological change, the process of digitization and digitalization have facilitated the formulation of data-intensive business models and provided a wide range of opportunities to create and adopt data-driven innovations by different economic agents.

The incremental growth of digital technologies supported by the information and communication technology (ICT) industry has enabled the expansion of the digital economy. As of now, digital transformation has influenced many aspects of modern life, including the way of conducting of business, productivity indicators, international trade, and has reconfigured economic structures and created new sources of consumer surplus. Following Katz et al. (2014) digitalization process reshapes the social and economic behavior by the extensive adoption of digital technologies, expanding the usage of platforms, online government services, electronic commerce, social networks and the access to online information.

Big-data, increasingly affordable computational power and faster connectivity are among the main elements that have paved the way to a data-driven economy, these technologies were adopted early by ICT related industries and large firms, bringing alongside data-driven innovations using technologies such as artificial intelligence, machine learning, cloud computing, automation, internet of things, data analytics, and data trading. Small and Medium Entreprises (SMEs) have had relatively limited access to these tools given the complexity and the type of internal capabilities that these technologies demand.

Margaras (2018) illustrates how northern Europe is leading the region as regards the development and usage of data-driven technologies. According to the European Commission, the ICT industry has grown 4.8% between 2006 and 2018, also it is expected to accelerate its growth even more during and after the COVID-19 pandemic. In the age of a pervasive digital economy, data-driven technologies developed and supplied mainly by the ICT sector, which provides a variety of products and services such as infrastructure, platforms, storage, business analytics software among others, contribute to handle the massive amount of data that is generated every second. In addition, ICT-intensive sectors such as healthcare, finance, education, energy, manufacturing, supply chain and many others, are demanding increasingly sophisticated digital solutions to run their operations and compete in a highly datafied environment (Zillner et al., 2016).

As the digital economy keep expanding, the incentives for companies to invest in digital tools increase. In the enlargement and deepening of the digital economy countries will face technological and knowledge challenges to develop a strong digitalization strategy that facilitates Data-Driven Innovation (DDI) (Bukht and Heeks, 2018). Over the years, the European Union (EU) has developed robust foundations for the advancement and implementation of digital technologies both in industry and in services. In June 2021, the EU has updated this policy to consolidate previous gains and enter a "*Digital Decade*", which aim to promote a digital society based on the core values of the EU and achieve a dynamic digital-driven economy by 2030. The execution of these multi-country projects will engage the objectives presented in the "digital compass": government, business, skills, and infrastructure. These policies not only target leading sectors such automobile manufacturing, transportation and aerospace equipment, chemical and pharmaceutical goods, software and electronics, but also traditionally less ICT intensive sectors like agroindustry, SMEs, startups and others potential ICT-users.

Despite the digital agenda that European countries have followed, there are still gaps in terms of connectivity and the use of digital technologies by different economic agents. The OECD report "*OECD Digital Economy Outlook 2020*" reveals that there are significant gaps among firm sizes in terms of the adoption of digital technologies specially in big-data analytics, artificial intelligence (AI) and cloud computing. Furthermore, these gaps become more evident if compared with smaller firms located in less developed European subregions. OECD (2017) describes how the absence of a strong and developed ICT sector hinders the data revolution. The lack of awareness, shortage of skilled human capital and the absence financial leverage are some of barriers that are facing SMEs in this digital transformation OECD (2020).

The aim of this research is to model and predict the achievement of DDI within a particular SMEs that are classified as (from now on) ICT-user[1] industries in Europe, by running a machine learning (ML) algorithm (regression trees, bagged trees and random forest, to name a few). The possible sources of micro-data are the European Patent Office (EPO), the Community Innovation Survey (CIS) and the Community Survey on ICT Usage and E-Commerce in Enterprises will be used depending on restrictions of data protection laws. Making special emphasis on the possible technological spillovers that could create the usage of ICT tools in the adoptions of DDI by the ICT-user industries. The quantitative analysis of this research project will benefit from the most recent methods, conceptual frameworks and theories developed in the digital economics and economics of innovation literature.

The rise of digital technologies has brought new elements that enable companies to develop new forms of innovation, this research project will focus on the type of innovation that is driven by data. First, considering the different aspects of DDI found in the literature Compagnucci et al. (2018), Babu et al. (2021), OECD (2017), Trabucchi and Buganza (2019), Zillner et al. (2016), the definition of DDI is the utilization and exploitation of any form of dataflows through data-analytics tools in the method to create value, new processes, products, and services. With the advent of the so-called fourth industrial revolution, data has become the main axis of value creation, the diffusion and development of different enabling technologies in the field of data management showed an average annual growth

[1] SMEs that pursue the use of ICT tools to develop their businesses.

of 22.5% since 2010 (EPO, 2020). Second, artificial intelligence in a narrow definition refers to the capability of machines to self-program using flows of new data. Nowadays, "*AI is able to collect and process signals via sensors or other methods; classify, learn, reason and predict possible outcomes; interact with people or other machines*" (Compagnucci et al., 2018), AI is a technology that is considered of high economic potential and estimated impact, the International Data Corporation (IDC) estimates that the worldwide revenues on AI in 2021 may amount to $327.5 billion. Considering the different scenarios where data has become the new fuel that drives these new digital technologies (Agrawal et al., 2018), evidence has shown that the use of big data becomes an important determinant in the propensity to innovate among firms (Niebel et al., 2019).

Early studies in the field of big data in different disciplines has permitted new insights into how to exploit data to create and capture value. In the field of administrative sciences Trabucchi and Buganza (2019) point out that scholars began to focus on the field of management where companies found an effective way to make more accurate decisions through big data analytics. In addition, the adoption of these techniques is gaining momentum in related fields such as economics, marketing, supply chain, innovation, finance and others.

The following research questions guide this inquiry: what is the role of connectivity and usage of digital tools in the process of DDI? According to the decision tree model, what are the decision paths that ICT-user industries take to develop DDI? What would be the main digital enablers of DDI for these particular SMEs in Europe? How the adoption of digital technologies influences the competitive advantage of these SMEs?

The proposed approach to the subject of DDI within the ICT-user industries involves a comprehensive literature review, in order to identify the theoretical framework and the state of the art, then, identify the mainstream arguments, techniques and methods found in the literature and evaluate the main avenues for further extensive research. The approach of such sophisticated topics requires a careful review of the prediction techniques that Machine Learning (ML) offers.

The European Commission has recently launched the Digital Europe Program that establish priorities on bringing digital tools and technologies to businesses, citizens, and public administrators. To tackle the existing digital gaps at different levels, programs like Horizon Europe and Connecting Europe Facility Digital have also joined to the Digital Europe Program. Beyond the connectivity gap, the digital agenda of these programs aim to develop national strategies, solutions and build capabilities in areas such as high-performance computing, cybersecurity; AI for economic growth; advanced digital skills; increase productivity; promote a democratic and sustainable digital society with a fair competitive digital economy. It is deemed of key importance the achievement of successful capacity-building strategies by means of diffusion and adoption of these initiatives among ICT-user firms and sectors.

Analyzing the global current events where the COVID-19 pandemic has exposed different challenges that brings the digitalization process, Davenport and Bean (2021) revealed that COVID-19 has caused a decline on data-driven initiatives for small firms. Consequently, it is pertinent to understand how these digital transformations will strengthen the ICT industries and how the government will conduct the strategies to build the digital capabilities for SMEs that are ICT-user industries to create value through data.

Beyond the incentives created by the new digital policy promoted by the European Union, where entrepreneurs, government and citizens will have opportunities to develop and adopt digital tools in order to create value, it will also be an opportunity for academia to explore, experiment, conceptualize, explain and contribute to the study of the opportunities and challenges that Data-Driven innovation may pose to society.

OBJECTIVES

- Identify the different factors that influence DDI development in the SMEs classified as ICT-user industries.
- Determine the influence of big-data in the process of DDI for the SME that are ICT-user.
- Analyze ML techniques related to decision trees models and the applications in the economic field.
- Compare the results obtained with previous research to identify limitations, advantages and disadvantages of the methodologies applied.

- Contribute to the understanding of how digital technologies may enable Data-Driven Innovations in European SMEs.

METHODOLOGY AND EXPECTED RESULTS

Machine learning is the discipline where computers are programmed to have the ability to learn from data (Géron, 2019). ML has a variety of powerful prediction techniques that are widely used by data scientist, computer engineers and more recently by economists (Athey and Imbens, 2019; Varian, 2014; Woloszko, 2020) to solve either classification or regression problems. Following Woloszko (2020) and Varian (2014), ML techniques have different ways to approach and model complex relationships. Algorithms such as decision trees, bagged trees, random forest, gradient boosting trees and others can easily capture non-linear patterns, economic complexity, counterfactual effects and structural changes. This research will focus on the *supervised learning* method which usually has three distinguishable stages after splitting the data into independent sets for the purpose of training, testing and validation. First, the algorithm will learn from the training set to estimate the model. Second, the process of validation will evaluate and choose the model. And third, the testing set will determine the performance of the selected model.

Despite the advantages of using the nonparametric prediction methods offered by ML, there are still some improvements that scholars have been able to perform in terms of causal inference in the field of social sciences. Athey (2019), (2015); Athey and Imbens (2019) and Wager and Athey (2018) have been able to develop different techniques that help to capture causality and estimation of optimal policy within tree-based models. Other studies by Goulet Coulombe et al. (2019), Guerzoni et al. (2020), Woloszko (2020) not only contributed to the integration of econometrics and ML but also suggested to keep delving on how ML may bring relevant elements in the econometric toolbox.

Using the CIS and EPO databases the ML model may predict whether ICT-user industries develop DDI or not, including different independent variables such location, company size, investment in R&D, technology resources only to name a few. The CIS database is the main source of data to measure innovation activities in Europe according to the definitions of the Oslo Manual. Furthermore, the EPO database incorporate more than 120 million patent documents with more than 250 thousand classification entries. Within this classification entries, patents related to methods or arrangements for data processing are classified under the code "G06F" according to the Cooperative Patent Classification. Choosing the best ML technique depends on the structure and the quality of the dataset. On the other hand, using the Community Survey on ICT Usage and E-Commerce in Enterprises database using the same ML techniques will be possible to measure the impact of digital tools (such as ICT specialist and skills, the use of cloud computing services, e-commerce, big data analysis, internet of things, access and use of the internet, among others) over ICT-user industries.

The results of this project may provide insights related to the determinants that influence DDI among the ICT-user industries in Europe according to their financial, technological and knowledge endowments as well as the effects of the adoptions of digital tools and may help to identify optimal adoption patterns for SMEs. The Digital Decade seems to be an interesting and promising program to achieve competitiveness and economic dynamism, but it will require the active participation of research initiatives that inform both business leaders and decision makers on the most effective paths towards Data Driven Innovation.

A successful adoption of data driven initiatives promises a deeper level of digital transformation necessary to remain competitive within the digital economy. However, not all companies (especially small firms) have the financial leverage and specialized personnel to exploit digital tools and create value through data analytics and thus pave the way to a digital business. The results not only will allow to map main characteristics and digital capabilities built by the ICT-user industries, but also to evaluate and formulate potential improvements in the overarching digital policy that Europe is planning to implement.

WORK PLAN

DATA-DRIVEN INNOVATION: LEVERAGING DIGITAL TRANSFORMATION FOR COMPETITIVENESS IN ICT-USER SECTORS	2021	2022				2023				2024			
	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
1. Literature review													
Bibliometric or topic modeling approach	■	■											
Concept mapping		■											
Identify theoretical approaches and frameworks		■											
2. Definition of Scope and Reach													
Research limitations			■										
Drafting of reach				■									
3. Conduct research													
Data Collection and Analysis					■	■							
4. Writing Process													
Chapter A							■						
Chapter B								■					
Chapter C									■	■			
Collection of observations and findings											■	■	
Drafting of final document												■	
Proofreading and editing												■	
5. Submission and defense													
Submission													■
Defense													■
6. Meeting Thesis Supervisor	●	●	●	●	●	●	●	●	●	●	●	●	●
7. Participation in Conferences													
Conferences and Simposia				■				■				■	

REFERENCES

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines : the simple economics of artificial intelligence*. Harvard Business Press.

Athey, S. (2019). The Impact of Machine Learning on Economics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 507–5049). University of Chicago Press.

Athey, S. (2015). *Machine Learning and Causal Inference for Policy Evaluation*. 5–6. https://doi.org/10.1145/2783258.2785466

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know about. *Annual Review of Economics*, *11*, 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Babu, M. M., Rahman, M., Alam, A., & Dey, B. L. (2021). Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms. *Annals of Operations Research*, *0123456789*. https://doi.org/10.1007/s10479-021-04077-1

Barham, H. (2017). Achieving competitive advantage through big data: A literature review. *PICMET 2017 - Portland International Conference on Management of Engineering and Technology: Technology Management for the Interconnected World, Proceedings*, *2017-Janua*(August), 1–7. https://doi.org/10.23919/PICMET.2017.8125459

Bukht, R., & Heeks, R. (2018). Defining, conceptualising and measuring the digital economy. *International Organisations Research Journal*, *13*(2), 143–172. https://doi.org/10.17323/1996-7845-2018-02-07

Compagnucci, S., Della Porta, M. R., & Massaro, G. (2018). *DATA-DRIVEN INNOVATION & Artificial Intelligence which strategy for Europe?* . www.i-com.it

Davenport, T. H., & Bean, R. (2021). *Big Data and AI Executive Survey 2021*. 1–15. https://c6abb8db-514c-4f5b-b5a1-fc710f1e464e.filesusr.com/ugd/e5361a_76709448ddc6490981f0cbea42d51508.pdf

EPO. (2020). Patents and the Fourth Industrial Revolution. *European Patent Office*. http://documents.epo.org/projects/babylon/eponet.nsf/0/06E4D8F7A2D6C2E1C125863900517B88/$File/patents_and_the_fourth_industrial_revolution_study_2020_en.pdf

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc. https://doi.org/10.1201/9780367816377

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2019). *How is Machine Learning Useful for Macroeconomic Forecasting? **. https://economics.sas.upenn.edu/system/files/2019-03/GCLSS_MC_MacroFcst.pdf

Guerzoni, M., Nava, C. R., Nuccio, M., Guerzoni, M., Nava, C. R., Start-ups, M. N., Guerzoni, M., & Nava, C. R. (2020). Start-ups survival through a crisis . Combining machine learning with econometrics to measure innovation with econometrics to measure innovation. *Economics of Innovation and New Technology*, *0*(0), 1–26. https://doi.org/10.1080/10438599.2020.1769810

Katz, R., Koutroumpis, P., & Callorda, F. M. (2014). Using a digitization index to measure the economic and social impact of digital agendas. *Info*, *16*(1), 32–44. https://doi.org/10.1108/info-10-2013-0051

Margaras, V. (2018). The regions in the Digital Single Market – ICT and digital opportunities for regions and cities. In *European Parliament Research Service* (Issue April).

Niebel, T., Rasel, F., & Viete, S. (2019). BIG data–BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, *28*(3), 296–316. https://doi.org/10.1080/10438599.2018.1493075

OECD. (2017). *Development Co-operation Report 2017. Data for Development*. OECD. https://doi.org/10.1787/dcr-2017-en

OECD. (2020). OECD Digital Economy Outlook 2020. In *OECD Digital Economy Outlook 2020*. OECD. https://doi.org/10.1787/bb167041-en

Trabucchi, D., & Buganza, T. (2019). Data-driven innovation: switching the perspective on Big Data. *European Journal of Innovation Management*, *22*(1), 23–40. https://doi.org/10.1108/EJIM-01-2018-0017

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspective*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Woloszko, B. N. (2020). *Adaptive Trees : A New Approach To Economic Forecasting* (Issue Working papers no. 1593).

Zillner, S., Becker, T., Munné, R., Hussain, K., Rusitschka, S., Lippell, H., Curry, E., & Ojo, A. (2016). Big data-driven innovation in industrial sectors. In *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_9