8-31-2020

Mind maps and machine learning: An automation framework for qualitative research in entrepreneurship education

Yasser Farha
New Jersey Institute of Technology

ABSTRACT

MIND MAPS AND MACHINE LEARNING: AN AUTOMATION FRAMEWORK FOR QUALITATIVE RESEARCH IN ENTREPRENEURSHIP EDUCATION

by
Yasser Farha

Entrepreneurship Education researchers often measure entrepreneurial motivation of college students. It is important for stakeholders, such as policymakers and educators, to assert if entrepreneurship education can encourage students to become entrepreneurs, as well as to understand factors that influence entrepreneurial motivation. For that purpose, researchers have used different methods and instruments to measure students' entrepreneurial motivation. Most of these methods are quantitative, e.g., closed-ended surveys, whereas qualitative methods, e.g., open-ended surveys, are rarely used.

Mind maps are an attractive qualitative survey tool because they capture the individual's reflections, thoughts, and experiences. For Entrepreneurship Education, mind maps can be utilized to measure students' entrepreneurial motivation. However, qualitative analysis of mind maps in business studies has been manually performed through human coding, which is time-consuming and labor-intensive, and of questionable reliability when more than one person does the analysis.

This dissertation provides a novel automation framework to address these challenges with an interdisciplinary solution that integrates deductive and inductive qualitative content analysis approaches with the computational power of machine learning algorithms and statistical Natural Language Processing to automate the analysis of mind maps. The framework includes four sequential steps: selecting a qualitative content

analysis approach, collecting and preprocessing mind maps, automating the analysis, and validating the reliability and model evaluation.

Experimentation and hypotheses testing for the automation framework are performed. The results show that the performance of classification models when applied to the automated deductive analysis of mind maps, and the performance of Structural Topic Model when applied to the automated inductive analysis of mind maps, are similar to that of manual mind maps analysis.

The utility of mind map topology in the process automation is evaluated. Findings indicate that even though inserting mind map topology as features into the dataset positively affects performance, the improvement is not statistically significant. On the other hand, treating nodes as the unit of analysis while applying Structural Topic Model to automate inductive analysis generates latent topics that follow a similar pattern to manual analysis.

This study examines the feasibility of applying the automation framework to Entrepreneurship Education research. Text classification algorithms and STM are used for the first time to automate the analysis of mind maps, and STM is applied for the first time in Entrepreneurship Education research.

The automation framework offers a unique and advanced qualitative research design that can be employed by EE researchers to benefit the EE best practices. The automation framework can enhance EE qualitative research by extracting textual statistical inference, shortening labor and time required by the analysis, and measuring entrepreneurial motivations with machine learning and Natural Language Processing techniques.

MIND MAPS AND MACHINE LEARNING: AN AUTOMATION FRAMEWORK FOR QUALITATIVE RESEARCH IN ENTREPRENEURSHIP EDUCATION

by

Yasser Farha

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Business Data Science

Martin Tuchman School of Management

August 2020

APPROVAL PAGE

MIND MAPS AND MACHINE LEARNING: AN AUTOMATION FRAMEWORK FOR QUALITATIVE RESEARCH IN ENTREPRENEURSHIP EDUCATION

Yasser Farha

Dr. Cesar Bandera, Dissertation Advisor Date
Associate Professor of Entrepreneurship, Martin Tuchman School of Management, NJIT

Dr. Zhi Wei, Committee Member Date
Professor of Computer Science, Ying Wu College of Computing, NJIT

Dr. Michael Ehrlich, Committee Member Date
Associate Professor of Finance, Martin Tuchman School of Management, NJIT

Dr. Stephen Taylor, Committee Member Date
Assistant Professor of Research Methods, Martin Tuchman School of Management, NJIT

Dr. Eric Liguori, Committee Member Date
Professor of Entrepreneurship, Rohrer College of Business, Rowan University

BIOGRAPHICAL SKETCH

Author: Yasser Farha

Degree: Doctor of Philosophy

Date: August 2020

Undergraduate and Graduate Education:

- Doctor of Philosophy in Business Data Science,
  New Jersey Institute of Technology, Newark, NJ, 2020

- Master of Science in International Business,
  Florida International University, Miami, FL, 2014

- Bachelor of Science in Business Administration,
  Lynn University, Boca Raton, FL, 2013

Major: Business Data Science

Presentations and Publications:

Farha, Y., & Bandera, C. "Assessing Student's Entrepreneurial Disposition by Text Classification Models," Graduate Research Day, Newark, NJ, November 2019.

Farha, Y., & Bandera, C. "Lexicon-Based Text Analysis Model for Entrepreneurship Education," The First Leir Research Institute Conference, Newark, NJ, April 2019.

Farha, Y., & Bandera, C. "Automated Text Analysis Models for Entrepreneurship Education," Proceedings of the United State Association for Small Business and Entrepreneurship Conference, St. Pete Beach, FL, January 2019.

This dissertation is dedicated to my parents, wife, children, and advisor. A special feeling of gratitude to my loving parents; my father and mother, who have encouraged, supported and motivated me throughout my educational journey. I will always be grateful for their unconditional support, guidance, and love.

I dedicate this dissertation to my wife, Nesren, who has always been supportive and never left my side throughout the process. She took care of my children while I was busy studying and attending courses. I will always be thankful for all she has done.

I dedicate this dissertation to my two angels, Teema and Hatim, who have changed my life for the better and pushed me to achieve my goals.

I dedicate this work to my research advisor Dr. Cesar Bandera for being more than an advisor. Since our first meeting, he has always demonstrated enthusiasm and appreciation for my efforts, he listened carefully to what I said and then provided me with insightful feedback. I will be forever indebted to him.

ACKNOWLEDGMENT

I would like first to acknowledge and express my deepest gratitude to the Almighty Allah for His blessings and grace. To Him, I owe everything.

I wish to acknowledge and especially thank Dr. Cesar Bandera, my committee chair, for his countless hours of inspiring, supporting, reading, reflecting, and ,most of all, being there when needed. I was blessed working under his supervision, he has always been very generous and kind to me, and words cannot describe my gratitude to him.

In addition, I would like to acknowledge and deeply thank my committee members who have generously given their precious time and expertise to better my work. I thank them for their contribution and their unlimited support. Thank you, Dr. Zhi Wei, Dr. Michael Ehrlich, Dr. Stephen Taylor, and Dr. Eric Liguori, for agreeing to serve on my committee.

I also wish to acknowledge and be thankful to my country Saudi Arabia for its financial support; they have put their faith in me and in the investment of my education.

I must acknowledge the following people as well, who have assisted me throughout my doctorate program: Dr. Cheickna Sylla, Dr. Yi Chen, and Dr. Dantong Yu. Also, I would like to thank my school for giving me the opportunity to conduct my research and providing any assistance requested.

Finally, I am grateful to my parents, wife, and children, whose love, support, and prayers are with me in whatever I pursue; they have always provided me with a constant source of inspiration and strength.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF DEFINITIONS

Accuracy	How closely an instrument measures the true or actual value of the process variable being measured or sensed.
Corpus	A large and structured collection of documents
Label	What a text is assigned to, e.g., a set of classes includes three labels.
Node	A branch within a mind map that may contain texts, numbers, or letters.
Unit	In sampling, it means an object, e.g., a person. In text analysis, it means the identified chunk of texts, e.g., a document.
Token	The smallest parts of texts, such as words, pronouns, and numbers
Vector	It refers to a vector space, so a four components vector has four dimensions in the vector space

CHAPTER 1

INTRODUCTION

1.1 Objectives

This dissertation consists of two general objectives. First, design and develop an automation framework for the qualitative studies in Entrepreneurship Education (EE). This general objective includes three specific objectives:

- Utilize mind maps as a qualitative data collection tool.
- Apply machine learning techniques to automate the Qualitative Content Analysis (QCA) of mind maps.
- Exploit mind map topology in the process automation.

The second general objective is to apply the automation framework, as a qualitative research design, into two EE problems to test its feasibility. First, classification models are applied to automate deductive analysis of mind maps collected from college students of two different cultures, U.S. and France, but taking a similar entrepreneurship course. Second, Structural Topic Model is applied to automate inductive analysis of mind maps collected from students of two different academic colleges, business and computer science, but enrolling in the same entrepreneurship course.

Section 1.2 presents background information of the motivations behind this research.

1.2 Background Information

Entrepreneurship Education (EE) research is undertaken for many reasons and serving a wide variety of goals (Blenker et al., 2014; Pittaway & Cope, 2007). Blenker et al. (2014) condense the purposes of EE research into two categories. The first category includes measuring the impact of EE to test and ensure its viability (Oosterbeek, Van Praag, & Ijsselstein, 2010). The second category includes inspecting the mechanisms and dynamics of teaching entrepreneurship (Neck, Greene, & Brush, 2014), the communication of entrepreneurship courses, programs, and best practices (Liguori et al., 2018).

As part of the first research category of EE, researchers measure the impact of EE on college students, e.g., (Farhangmehr, Gonçalves, & Sarmento, 2016; Oosterbeek, Van Praag, & Ijsselstein, 2010; Vanevenhoven & Liguori, 2013; Vesper & Gartner, 1997). The impact of EE research is defined as "*the majority of studies that analyze the impact of entrepreneurship education on entrepreneurial attitudes, intentions, and venture activities*" (Lorz, Mueller, & Volery, 2013, p. 123).

The attention paid by EE researchers to measure the impact of EE on students is justified (Pittaway & Cope, 2007). First, policymakers and educators need to know if EE can induce college students to become entrepreneurs because there are growing investment and expansion of EE programs and courses within colleges of business as well as other colleges (Curth et al., 2015; Souitaris, Zerbinati, & Al-Laham, 2007). Second, EE impact seeks to investigate variables that may be influential to such an impact, such as curriculum and method of teaching (Liguori et al., 2018), academic majors (Kolvereid & Moen, 1997), and cultural norms (Bandera et al., 2018a).

Entrepreneurship Education impact studies are essential for the improvement and development of entrepreneurship as an academic discipline because they assist educators in increasing the chances of producing optimal EE impacts on students (Duval-Couetil, 2013). For example, these studies' findings can help educators mitigate mediator factors of the EE impact by tailoring specialized EE courses and programs among different cultures and academic majors (Katz, 2003; Maresch et al., 2016).

1.2.1 Measurement in Entrepreneurship Education Research

Researchers and educators in EE have assessed the impact of EE using a variety of measurements that serve as an indication of entrepreneurial outcomes, including entrepreneurial intention (Bae et al., 2014) and entrepreneurial motivation (Giacomin et al., 2011).

Entrepreneurial motivation refers to a set of personal goals, whether instrumental or final (Nuttin, Lorion, & Dumas, 1984) that can be achieved through entrepreneurship (Farhangmehr, Gonçalves, & Sarmento, 2016). The use of entrepreneurial motivation has gained moderate attention in the field of entrepreneurship. Carsrud and Brännback (2011) claim that the use of entrepreneurial motivation has been ignored, and they ask for renewing interest in investigating entrepreneurial motivation, which they argue to be critical and directly interrelated to entrepreneurial cognitions, intentions, and behavior. They believe that entrepreneurial motivation is an important topic that still requires more study (Carsrud & Brännback, 2011).

Shane, Locke, and Collins (2003) believe that considering the motivations behind making entrepreneurial decisions is necessary to develop entrepreneurship theory. They review several characteristics that have been found to affect the entrepreneurial process and introduce them as motivations, such as need for achievement, risk taking, locus of control, and self-efficacy.

In the context of EE impact studies, entrepreneurial motivation is treated as an outcome measure for a college course or program of study (Blenker et al., 2014). However, some works indicate that much of the research in the field has not provided indisputable empirical evidence for the perspective that entrepreneurship education increases entrepreneurial motivation, e.g., (Duval-Couetil, 2013; Farhangmehr, Gonçalves, & Sarmento, 2016; Martin, McNally, & Kay, 2013).

1.2.2 Methods in Entrepreneurship Education Research

For the assessment of college students' entrepreneurial motivation, different research methods have been utilized. For example, by using a closed-ended survey as an instrument for their study, Hsu, Shinnar, and Powell (2014) have demonstrated that entrepreneurship education can increase students' entrepreneurial motivation and intentions. The survey includes questions constructed on the concepts of entrepreneurial self-efficacy, desirability, intentions, and other related outcomes, including need for achievement, financial security, making money, and being independent.

Basu and Virick (2008) have utilized a pencil and paper to measure students' intentions to become entrepreneurs. Their survey consists of nine items to measure attitude toward entrepreneurship and are scaled on a range of five multiple choices. The questions have been adapted from (Kolvereid, 1996). Moreover, many other studies including (Barakat, Boddington, & Vyakarnam, 2014; Duval-Couetil, Reed-Rhoads, & Haghighi, 2012; Franco, Haase, & Lautenschläger, 2010; Hsu, Shinnar, & Powell, 2014; Kim-Soon, Ahmad, & Ibrahim, 2016; Liu et al., 2019 D Zhao, 2019; Pihie & Bagheri, 2011; Solesvik,

2013) have also used closed-ended survey instruments, which represent a quantitative method, to measure students' entrepreneurial motivations and intentions.

The observation of the frequent use of quantitative methods corresponds to one of the significant methodological criticisms that EE impact studies have recently received, which is the shortage of applying qualitative methods. Lorz, Mueller, and Volery (2013) present a systematic literature review to analyze the methods used in EE impact studies. They report that only two out of thirty-nine reviewed studies have implemented a qualitative method. Thus, they conclude that impact studies suffer a lack of using qualitative methods.

Research in entrepreneurship, in general, has shown a similar lack of applying qualitative methods. Mullen, Budeva, and Doney (2009) have analyzed the research methodologies used by entrepreneurship researchers between January 2001 and February of 2008. The analysis includes 665 papers published in the Journal of Small Business Management, Journal of Business Venturing, and Entrepreneurship Theory and Practice, in which 478 are sorted as empirical papers. From the 478 empirical papers, only 50 are qualitative (case studies, interviews, and observations), about 10%, while 428 empirical studies are quantitative, about 90%. Most of the quantitative studies relied on data gathered from surveys.

Therefore, several scholars have indicated the importance and need for qualitative methods in future studies, e.g., (Gartner & Birley, 2002; Lorz, Mueller, & Volery, 2013; Molina-Azorín et al., 2012; Mullen, Budeva, & Doney, 2009).

1.2.3 Importance of Qualitative Methods

Blenker et al. (2014) report that despite generalizability and comparability achieved by quantitative methods, they suffer limited estimation of the variance of the education they attempt to measure. Rahman (2017) indicates that quantitative research methods tend to overlook participants' experiences as well as what they mean by something and they only grasp a snapshot of a phenomenon.

Although qualitative methods suffer from limited generalizability, qualitative methods are full of contextually sensitive explanations and best pedagogical practices (Blenker et al., 2014). Besides, Rahman (2017) finds that qualitative methods possess some strengths for language evaluation and testing, such as obtaining more profound insights into designing, administering, and interpreting related evaluation and testing. Despite some weaknesses of qualitative methods, such as time-consuming analysis and smaller sample sizes, qualitative methods can explore participant's understanding, feelings, perceptions, and behavior (Patton, 2014; Rahman, 2017).

The advantages of qualitative methods are neglected in EE impact studies (Lorz, Mueller, & Volery, 2013). EE research needs more qualitative studies, not only because there has been insufficient use but also because of the several advantages that might be achieved. Lorz, Mueller, and Volery (2013) and Zahra (2007) suggest that EE researchers should pay more attention to the context of their research because importing existing theories from other disciplines can overlook significant findings. By understanding the nature, active, and richness of their studies' context through qualitative methods, researchers can provide more insightful explanations and findings that can facilitate students learning of entrepreneurship.

1.2.4 Qualitative Research Instruments

Principal data collection instruments in EE qualitative research include interviews, open-ended surveys, and focused discussion (Blenker et al., 2014; Mullen, Budeva, & Doney, 2009). Interviews are the most common qualitative data collection instruments in EE research, and open-ended surveys come second (Blenker et al., 2014).

Open-ended surveys can provide a direct view of a respondent's thinking (Roberts et al., 2014). For instance, RePass (1971, p. 391) contends that open-ended questions query viewpoints that "*are on the respondent's mind at the time of the interview,*" which they were presumably salient before answering the question and remain so afterward. Comparably, Iyengar (1996, p. 64) writes that open-ended questions have the advantage of "nonreactivity," which is different from closed-ended questions, "*open-ended questions do not cue respondents to think of particular causes or treatments.*" In EE setting and research, open-ended instruments can offer a deeper understanding of students' thinking because it allows them to express themselves without following specific and predefined options.

Participants in open-ended surveys nevertheless might not respond to some questions due to an inability to express one mind. Non-response to open-ended questions may stem from ineloquence rather than indifference, or because subjects lack the necessary rhetorical device (Geer, 1988). Therefore, as a unique alternative, mind maps can be utilized as an open-ended survey tool for future qualitative studies of EE impact. Haddock and Zanna (1998, p. 147) consider mind maps as an open-ended method "designed to allow the researcher to understand the responses individuals spontaneously associate with an attitude object, as well as how this information is organized in memory."

1.2.5 Mind Maps

Mind mapping is a technique in which an individual's perceptions, reflections, and experiences are represented visually by connecting ideas and concepts to a central subject (Buzan & Buzan, 2006). They are a diagrammatic representation of ideas and thoughts. They are less structured than concept maps and thus easier to generate. The process of drawing a mind map starts by placing the primary concern in the center node (Buzan & Buzan, 2006). From that center, lines are drawn out to make branches, with each branch consisting of words, and these branches can be subdivided into other branches to continue building the map outward (Buzan & Buzan, 2006). It is a process of generating ideas and concepts as a connected flow of thoughts. Figure 1.1 shows an example of a mind map drawn by a student as a response to the central subject of what motivates her to create a new venture.

Mind maps capture how participants organize and apply knowledge without constraints (Buzan & Buzan, 2006). This quality is viewed as a valuable element when assessing concerns about complex topics and reasoning (Somers et al., 2014). Moreover, the free form graphical nature of mind maps has been found to engage students more than regular written tasks. They also yield enhanced learning and attitudes across students of diversified learning preferences and cultures (Horton et al., 1993; Jones et al., 2012).

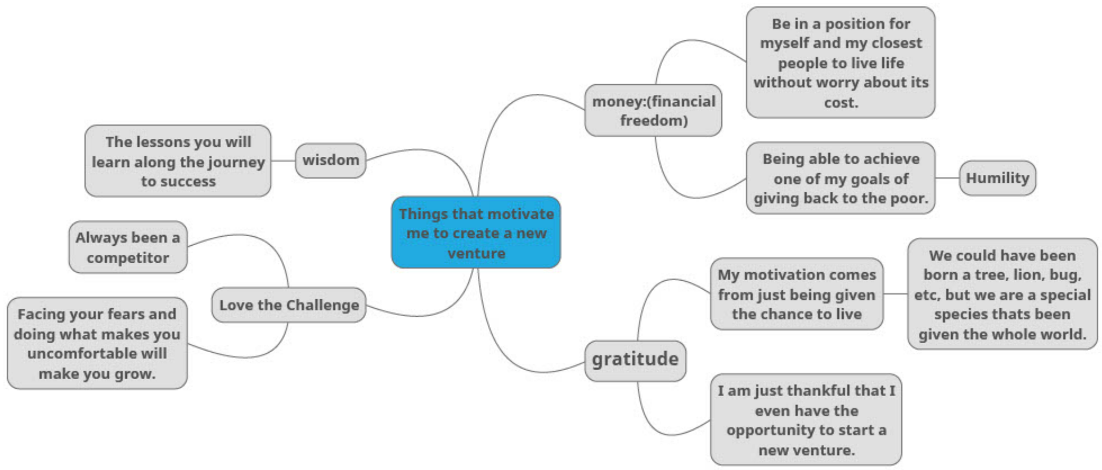

Figure 1.1 A sample of mind map drawn by a student to answer "Things that motivate me to create a new venture."

For these reasons, mind maps have been suggested to provide a unique tool for collecting individualistic and personalized data from research participants (Tattersall, Watts, & Vernon, 2007). They offer distinctive aspects to collect qualitative data. First, they start with a single theme or concern of which other ideas and concepts follow seamlessly (Buzan & Buzan, 2006). When participants are asked to complete mind maps, answers and concepts are connected, ordered, and presented as another means for researchers to collect data (Wheeldon, 2010). Second, they are free form, creative, and drawn by participants with easy to follow instructions (Wheeldon & Ahlberg, 2017).

Utilizing mind maps as a data collection tool in EE qualitative research instead of conventional tools, such as interviews, is more practical. Mind maps are assigned with minimal instruction by the researcher (Wheeldon & Ahlberg, 2017), providing a practical solution for the rigidity between quality and limited resources (Burgess-Allen & Owen-Smith, 2010), and facilitating the mission of qualitative analysis with participant-led segmentation of inputs (Tattersall, Watts, & Vernon, 2007).

1.2.6 Mind Maps Analysis

Different quantitative approaches have been used to analyze mind maps in social science research. For example, one approach compares outputs of mind maps to a "golden model" mind map; this called comparison with a criterion map (Jamieson, 2012). Ruiz-Primo, Schultz, and Shavelson (1997b) have introduced another technique "score map," analyzing mind maps quantitatively based on a graphic basis by counting nodes and edges. A third model is a hybrid model that merges the above two models (Beel & Langer, 2011). It includes comparing and counting of nodes and edges. Another approach assigns values to nodes connections based on the maps topology, such as their distance from the map's central subject, and then compute the sum of these link values as the score of a mind map (West et al., 2002).

None of these approaches apply qualitative methods to analyze data, for instance, the QCA method (Krippendorff, 2018; Mayring, 2014). For example, these approaches do not extract categories from within the textual data. Although they may estimate the overall depth and breadth of a mind map, they do not distinguish one semantic emphasis from another, such as in QCA.

On the other hand, there is a semantic analysis presented by Somers et al. (2014) and Bandera et al. (2018b), which scores mind maps by qualitatively coding nodes inside mind maps into categories. This approach enables a more in-depth analysis of nodes within a mind map. Somers et al. (2014) have used qualitative analysis to measure the differences in knowledge retention among business students. The authors have assigned the nodes in students' mind maps, with "successful business" as the central subject, into four groups, the disciplines of marketing, finance, human resource management, and strategy. Bandera et al. (2018b) have applied qualitative analysis of mind maps to distinguish cultural norms between U.S. and French entrepreneurship students.

1.3 Problem Statement

This dissertation addresses two main problems and related issues. The first main problem is the lack of using qualitative methods in EE impact studies, which is observed as a gap in the literature; the number of qualitative studies is remarkably low when compared to quantitative studies, e.g. (Gartner & Birley, 2002; Lorz, Mueller, & Volery, 2013; Molina-Azorín et al., 2012; Mullen, Budeva, & Doney, 2009). Consequently, EE research has not fully benefited from the advantages of qualitative methods, including acquiring contextually responsive descriptions and deeper insights into language assessment (Blenker et al., 2014; Rahman, 2017).

In addition to the lack of qualitative methods, Blenker et al. (2014) state that most of the reviewed quantitative and qualitative studies have only used descriptive data analysis techniques, while advanced techniques, such as regression analysis, have rarely been performed. The deployment of advanced data analysis techniques could advance EE qualitative research from mere description to an in-depth exploration of entrepreneurship impact.

The second main problem is that the qualitative analysis of mind maps in the business studies has been manually done through human coding (Bandera et al., 2018b; Somers et al., 2014), which is affected by a few critical issues. The manual analysis of mind maps is time-consuming and labor-intensive. The consistency of the analysis is also questionable, especially when the analysis is done by more than one analyzer, which may cause a reliability problem (Johnson, Penny, & Gordon, 2000).

The manual analysis may also neglect the potential contribution of mind map topology, e.g., the link between nodes, and other text richness features, such as word frequency. Although some quantitative and qualitative approaches to analyze mind maps have been manually used, Beel and Langer (2011) report that there is an insufficiency of analyses of mind maps, and information retrieval tasks have never been applied to mind maps. Therefore, the analysis of mind map content and structure needs more attention.

Mind maps can be more valuable as a qualitative data collection tool when having a robust qualitative analysis, exploiting their topology, and, most importantly, automating such process. The merger of rigorous qualitative analysis approaches with the computational power of machine learning techniques to analyze mind maps presents a novel methodology that researchers can utilize not only in the entrepreneurship or business fields but also in other academic disciplines.

1.4 Purpose Statement

The aim of this dissertation is to present a framework of machine learning techniques for the automated qualitative analysis of mind maps and the demonstration of this framework in the context of entrepreneurship education.

1.4.1 Automate Qualitative Content Analysis of Mind Maps

The qualitative analysis of mind maps in the business studies has been only performed manually through human coding of content (Bandera et al., 2018b; Somers et al., 2014). While mind maps possess several advantages over conventional qualitative data collection tools (Tattersall, Watts, & Vernon, 2007), they are still sharing the weaknesses of manual analysis, specifically, time-consuming and labor-intensive. These two problems, in addition to a reliability issue, require an advanced solution (Bandera et al., 2018a). Therefore, this research aims to automate QCA of mind maps with state-of-the-art machine learning and Natural Language Processing (NLP) techniques (Aggarwal & Zhai, 2012; Manning & Schütze, 1999).

The automation framework presented in this dissertation integrates the QCA approaches, deductive and inductive (Mayring, 2004), with the computational power of machine learning algorithms (Aggarwal & Zhai, 2012; Bishop, 2006) and statistical NLP techniques (Manning & Schütze, 1999) to automate the analysis of mind maps. Figure 1.2 shows a diagram of the integration.

The automation of mind maps analysis remarkably cuts both amounts of time and number of labors required by manual analysis. It also achieves other benefits of using computational techniques, such as retrieving information that is difficult for humans to detect, including counting frequent words and discovering latent topics.

1.4.2 Exploit Mind Map Topology in Automated Analysis

Mind maps offer a unique and smart thought visualization tool, which facilitates both participants' and analysts' tasks. Participants express themselves freely when using mind maps, and analysts find categories and clusters of data already segmented by participants (Wheeldon & Ahlberg, 2017; Wheeldon & Faubert, 2009). These advantages of mind maps being a data collection tool, should not be ignored when automating qualitative analysis. The link between nodes inside a mind map can be exploited in the process automation.

This research aims to examine the exploitation of mind map structure and topology in the automated analysis. This is vital because the topology of mind maps, as a distinct property of mind maps, is what distinguishes mind maps from any regular qualitative data collection tool.

1.4.3 Advance EE Research with an Automated Qualitative Analysis Framework

The need for EE research to qualitative studies and the benefits of them, especially for measuring students' entrepreneurial motivation, is a primary incentive behind this dissertation. The field demonstrates a shortage of qualitative methods when measuring EE's impact on college students (Gartner & Birley, 2002; Lorz, Mueller, & Volery, 2013; Molina-Azorín et al., 2012; Mullen, Budeva, & Doney, 2009). This dissertation aims to boost the number of qualitative studies in EE research by providing an automated framework that eliminates the barriers of using such methods. These barriers include time-consuming, labor-intensive, and possible inconsistent analysis (Johnson, Penny, & Gordon, 2000).

Besides the lack of using qualitative methods, other methodological concerns in EE research have been discussed, for example, the reliability and validity of analysis (Mullen, Budeva, & Doney, 2009) and the lack of performing advanced statistical analysis (Blenker et al., 2014). The automation framework also aims to mitigate such concerns. Regarding the validity and reliability of the analysis, the automation framework includes two prominent properties that emphasize robustness of qualitative analysis.

The framework follows precise guidelines for the qualitative analysis of mind maps (Hsieh & Shannon, 2005; Mayring, 2004, 2014; Prasad, 2008). These guidelines are planned out in step one of the automation framework, selection of QCA approach. Second, step four of the automation framework defines an established criterion for validity and reliability of the qualitative analysis (Egami et al., 2018), as well as presenting a variety of score metrics for model evaluation (Bradley, 1997).

For the advancement of statistical analysis, the automation framework allows analysts to perform advanced statistical models, such as regression analysis. The applications of EE problems, presented in Chapter 4, illustrate how the automation framework uses regression analysis to test case problems.

1.4.5 Measure Entrepreneurial Motivation with Machine Learning and Statistical NLP Techniques

There has been a call for using advanced technology in the research of EE. Mullen, Budeva, and Doney (2009, p. 287) state that "*a strong methodological foundation built on state-of-the-art research technologies is necessary to support further paradigmatic growth and maturation in the entrepreneurship research*." Entrepreneurial outcomes in general, and intentions and motivation in specific, have not been measured with a statistical NLP technique. This dissertation aims to measure entrepreneurial motivation for the first time with an NLP technique of topic modeling, precisely, Structural Topic Model. Structural Topic Model is applied to automate the inductive qualitative analysis of mind maps.

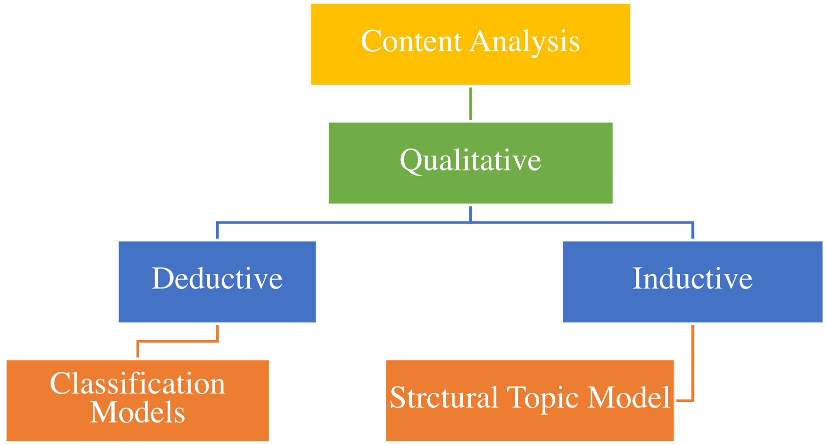

Figure 1.2 The integration of QCA and machine learning techniques, where the deductive approach is automated by text classification models, and the inductive approach is automated by Structural Topic Model.

1.5 Research Questions and Hypotheses

The study seeks to automate the process of QCA of mind maps used as a data collection tool by answering the following research questions:

- *RQ1a*: Can deductive qualitative content analysis of mind maps used as a data collection tool be automated?
- *RQ1b*: Can inductive qualitative content analysis of mind maps used as a data collection tool be automated?
- *RQ2a*: Can topology of mind maps be helpful to the process automation of deductive analysis?
- *RQ2b*: Can topology of mind maps be helpful to the process automation of inductive analysis?

1.5.1 Automation of Qualitative Content Analysis

This study assumes that QCA of mind maps can be automated by machine learning and NLP techniques. QCA includes two general approaches, deductive and inductive, which hold different processes. This research hypothesizes that a different technique automates each approach.

1.5.1.1 Automating Deductive Analysis. The manual deductive analysis is conceptually similar to supervised machine learning, specifically text classification, in that both require a reference or target for a classification task (Egami et al., 2018; Scharkow, 2013). A reference in text classification defines the possible values of output y for each data input x (Kotsiantis, Zaharakis, & Pintelas, 2007). The domain of y values is used as a target to supervise the learning process and build the mapping function f. For example, in text classification, a document, as input x, is classified to one of the pre-defined classes, as output y, of which the model learns and parameterizes the mapping function f. In the same manner, a reference in the deductive analysis guides manual coding by providing established classes or categories (Hsieh & Shannon, 2005).

Mayring (2014) argues that using advanced technology, such as text classification, for QCA is supported by three reasons. First, the textual material nowadays can be transferred into a software program. Second, QCA is a systematic, controlled, and step-by-step text analysis where the use of advanced technology tools can be beneficial. Third, in the last two decades, computational software has been developed, especially for qualitative text analysis. Within the context of the deductive approach, several studies have used classification models to automate deductive content analysis. For example, Scharkow (2013) used text classification to automate the content analysis of German online news. Scharkow concludes that supervised text classification is generally reliable for similar tasks. Preoţiuc-Pietro, Lampos, and Aletras (2015) applied text classification to predict occupational classes of Twitter users based on their profile content; their study's results confirmed the feasibility of using text classification to automate deductive content analysis.

For that reason, this research hypothesizes that the deductive content analysis of mind maps can be automated by applying text classification models; thus, the following hypothesis is formulated:

- *H_{1a}: The performance of automating the deductive analysis of mind maps with classification models is similar to that of human deductive analysis*

A computational experiment tests the above hypothesis. It is presented in the methodology chapter as "experiment number one," in which classification models automate the deductive analysis of mind maps.

1.5.1.2 Automating Inductive Analysis. The inductive approach of content analysis and unsupervised machine learning, such as topic modeling, extracts categories directly from text without using prior knowledge. In the inductive content analysis, analysts extract codes, categories, or themes directly from the text data without a reference (Hsieh & Shannon, 2005; Mayring, 2014). Topic modeling algorithms discover patterns from data without requiring labeled outcomes (Friedman, Hastie, & Tibshirani, 2001; Hinton, Sejnowski, & Poggio, 1999).

Baumer et al. (2017) have compared the results of analyzing text data between two methods: topic modeling and inductive analysis. They have concluded that both methods identify thematic patterns, grounded in the data, and require iterative process (Baumer et al., 2017). In addition to the methodological similarities, Baumer et al. (2017) have reported that results from both methods are at some level aligned. For analysis of open-ended surveys, Roberts et al. (2014) have found that Structural Topic Model, as a topic modeling, and human-coded analysis produce similar results. The authors have added that Structural Topic Model is helpful for survey researchers and experimentalists because it makes qualitative analysis of open-ended responses easier and more revealing.

This study hypothesizes that the inductive content analysis of mind maps can be automated by Structural Topic Model (STM) (Roberts et al., 2014). STM has an advantage over other topic models, such as Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003), because of the capacity to include documents' metadata, such as affiliation or treatment, in the analysis (Lucas et al., 2015; Roberts et al., 2013). For qualitative studies, STM enables observation of treatment effect or variables of interest with discovered latent topics, which leads to the possibility of testing a relationship between them, i.e., hypothesis testing (Roberts, Stewart, & Tingley, 2014).

Another major factor for using STM is the text size inside mind maps. Mind maps generate short texts. Beel and Langer (2011) analyze about 19 thousand mind maps and find that they have an average of 30 nodes per map and 4.8 words per node. In recent years, researchers have examined the use of topic modeling on short texts, e.g., (Cheng et al., 2014; Quan et al., 2015; Zuo, Zhao, & Xu, 2016), and conclude that specific topic models perform better than the traditional models when handling short texts, and one of those effective models is STM. Therefore, this research includes the following hypotheses:

- *H_{1b}: The performance of automating the inductive analysis of mind maps with Structural Topic Model is similar to that of human inductive analysis*
- *H_{1c}: Structural Topic Model outperforms Latent Dirichlet Allocation for automating the inductive analysis of mind maps*

A computational experiment tests the two hypotheses. The experiment is presented in the methodology chapter under "experiment number two," in which STM and LDA are applied to automate the inductive analysis of mind maps.

1.5.2 Mind Map Topology

Many studies explore the effectiveness of mind maps as a learning tool; nevertheless, a lack of methods to analyze mind maps has been reported (Beel & Langer, 2011). Besides the novelty of automating the qualitative analysis of mind maps, this dissertation assumes that the distinctive topology of mind maps can help automate qualitative analysis.

A mind map topology can be defined as a tree networking topology that "*starts out with a highest level or root level where a single node is connected to nodes in a second level of the hierarchy, second level nodes each connect to one or more nodes in third level, and each level fans out further*" (Sosinsky, 2009, p. 15). This definition implies that a mind map starts with one node as root, i.e., the central node, which like tree networking, connects to child nodes, which themselves can connect to grandchild nodes, and so forth, forming hierarchal levels of nodes.

Beel, Gipp, and Stiller (2009) posit that a link analysis among nodes of a mind map can be practical for information retrieval tasks, such as document summarization and document relatedness. For example, Figure 1.3 shows a mind map with five nodes in two branches, Beel, Gipp, and Stiller (2009) propose that node number 2 holds a high content relatedness to node number 3, but a lower content relatedness to node number 5 because it is under a different branch.

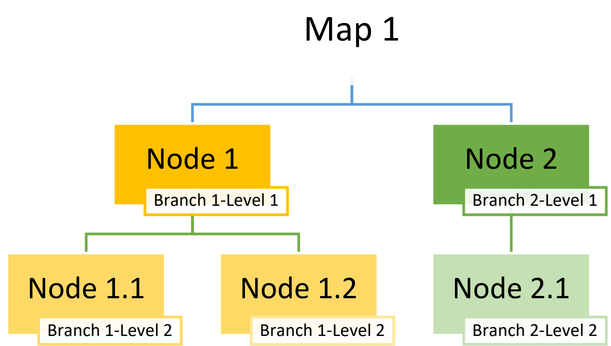

Figure 1.3 Five nodes inside one mind map.

In this study, this link among nodes inside a mind map can be utilized to enhance the process automation of deductive analysis. The automation of the deductive analysis of mind maps is a task of text classification, and feature extraction can play a crucial role in affecting such process (Ramya et al., 2017). Feature extraction is the generation of new features from the raw data (Guyon et al., 2008), which can assist text classification models to achieve better results (Humphreys & Wang, 2017). Thus, this research hypothesizes that using topology of mind maps, as a link analysis among nodes, to generate new features can improve the performance of automating deductive analysis of maps.

Two methods for generating such features are presented and tested in this paper. The methods examine two different perspectives for the relation between link analysis of nodes and classification tasks. The first method follows a similar path to the link analysis presented in (Beel, Gipp, & Stiller, 2009), which emphasizes the link analysis among nodes of one mind map based on their branch, hierarchal level, and the number of nodes in the same level. The first method is referred to as "General-link," because the feature extraction process does not distinguish nodes of different mind maps. Figure 1.4 shows the nodes features of two mind maps, yet, because maps have a similar topology, features of nodes are similar.

In contrast, the second method emphasizes link analysis among nodes of one mind map based on their branch, but each branch is distinct across all mind maps. This method is referred to as "Specific-link," because the extracted features differentiate nodes of distinct mind maps. Figure 1.5 shows the nodes' features of two mind maps with identical structure, but the extracted features are only similar for nodes belong to the same map and branch.

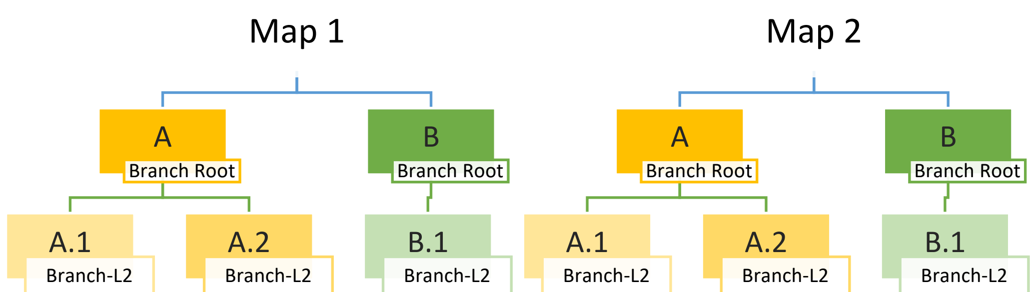

Figure 1.4 General-link analysis of two mind maps, where features "inside the boxes" for a given node are based on its branch and hierarchal level.

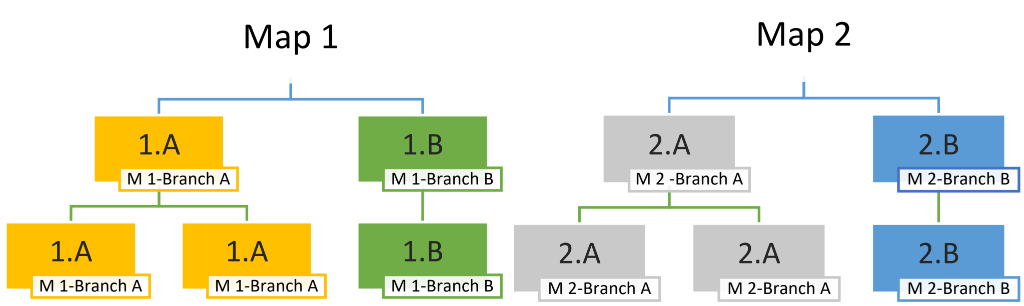

Figure 1.5 Specific-link analysis of two mind maps. Features extracted for a given node "inside the boxes" are determined by its map and branch. For example, nodes belong to one map and branch have the same features.

Thus, this research includes the following hypotheses, one for each method:

- *H_{2a}: Using topology of mind maps "General-link" as features improves classification performance when automating the deductive analysis of mind maps*
- *H_{2b}: Using topology of mind maps "Specific-link" as features improves classification performance when automating the deductive analysis of mind maps*

The two hypotheses are tested by comparing the performance of two classifiers "A" and "B". Classifier A uses mind maps' content, i.e., texts inside nodes, whereas classifier B utilizes the content of mind maps and topological features of those maps. A separate experiment tests each hypothesis.

Tattersall, Watts, and Vernon (2007) and Wheeldon and Ahlberg (2017) have asserted that mind maps facilitate the mission of qualitative analysis because of the participant-led segmentation of inputs, meaning that nodes are virtually drawn to represent one idea or concept. The human analysis of nodes in (Bandera et al., 2018b) signifies that assertion by assigning a node to one category representing the discussed concept in the node. Burgess-Allen and Owen-Smith (2010) studied over 18 thousand public mind maps and reported that the average words per node are 4.23.

This research hypothesizes that treating nodes as separate documents, which in the context of text analytics is defined as the unit of analysis (Aggarwal & Zhai, 2012), improves the performance of inductive process automation of STM by assigning a node to as close as one topic, which is commensurate to the human analysis of mind maps (Bandera et al., 2018b).

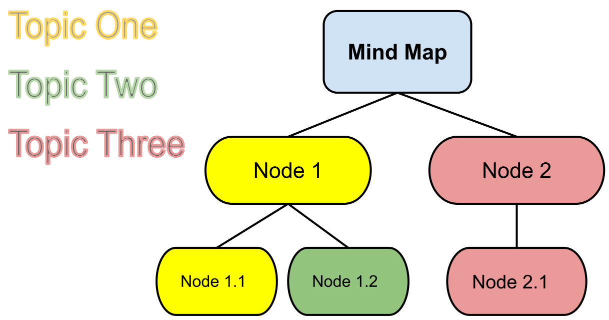

Figure 1.6 An example of three topics distribution over nodes inside one mind map.

Figure 1.6 shows an example of a mind map consisting of five nodes and three STM topics. A single topic dominates each node. Node 1 in Figure 1.6 may not be 100% topic one, because it is still proportionally distributed over the three topics, but since nodes are drawn to represent one concept, and they include an average of 4.23 words (Burgess-Allen & Owen-Smith, 2010; Tattersall, Watts, & Vernon, 2007; Wheeldon & Ahlberg, 2017), this research hypothesizes the following:

- *H_{2c}: Treating nodes as the unit of analysis improves STM performance by assigning nodes to a single "dominating" topic that exceeds the null probability of* $\hat{p} = .5$

This hypothesis is tested by computing the average of the dominant topic probabilities in all nodes, and check if this average is significantly larger than the value under the assumption of domination, $\hat{p} = .5$. The threshold of .5 indicates whether a probability of one topic inside a node is statistically more significant than the total remaining topics.

1.6 Significance of the Study

Entrepreneurship Education impact studies suffer a lack of using qualitative methods, in part because the analysis of such mechanisms is traditionally labor-intensive and time-consuming and thereby not scalable to extensive studies involving many students or regions. On the other hand, mind maps, which hold potential benefits for EE studies as a qualitative data collection tool (Bandera et al., 2018b), have only been manually analyzed, which as a process, suffers from the issues of labor-intensive and time-consuming.

The automation framework presented in this study addresses these challenges with an interdisciplinary solution that combines machine learning techniques, QCA approaches, and the constraints of open-ended survey mechanisms used in EE assessment. The automation framework provides EE researchers with a novel automated analysis for qualitative methods and utilizes mind maps as a data collection tool. The automation framework can also be used in other academic disciplines, such as communication studies, public health, and psychology.

This framework aims to automate the generation of analytics with which to improve the effectiveness of EE. The automation framework makes analyzing mind maps used in the EE research easier and scalable to large cohorts, more consistent and revealing, and capable of being used to evaluate differences among sample groups. It allows EE researchers to choose a QCA approach that suits the goal of their studies.

QCA is suitable to analyze mind maps collected in EE studies because it is a flexible research method (White & Marsh, 2006), describes data that requires some degree of interpretation (Schreier, 2012), and offers two distinct approaches to analyze data

(Mayring, 2004). The automation framework provides process automation for each of the two approaches, deductive and inductive.

Entrepreneurship Education research needs qualitative methods not only because there are few in the literature, but also because of the deep level of comprehension of context that might be achieved. Gartner and Birley (2002, p. 387) state that "*It is our opinion that many substantive issues in entrepreneurship are rarely addressed and that many of the important questions in entrepreneurship can only be asked through qualitative methods and approaches.*" The use of qualitative methods in EE studies encourages students to express their experiences in a way that quantitative methods cannot allow (Rahman, 2017). Students' entrepreneurial motivations can be heard directly instead of suggesting them to students.

Researchers and practitioners in the field of EE acknowledge the difficulty in assessing students' motivation to entrepreneurship, as well as measuring how some factors might affect that motivation (Morales-Gualdrón, Gutiérrez-Gracia, & Dobón, 2009). The automation framework opens the door for applying advanced qualitative methods that have not been applied to such assessments before.

The automation framework presented in this research can stimulate the use of qualitative methods in the field of EE because it allows for collecting substantial qualitative data with an innovative tool, mind maps, selecting which qualitative analysis to implement, i.e., deductive or inductive analysis, automating selected analysis with state-of-the-art machine learning techniques, classification models and STM, and validating the results. For instance, the use of mind maps is faster in collecting and analyzing data than standard qualitative tools (Burgess-Allen & Owen-Smith, 2010). With automating the mind maps analysis, the sample size of a qualitative study can be commensurate with that of a quantitative study, which can mitigate the generalizability and comparability issues of qualitative methods (Rahman, 2017).

The automation framework permits direct and advanced testing for qualitative data in EE. Two applications included in the experiments in Chapter 4 demonstrate the application of the automation framework to the statistical testing of EE case problems. For example, in the second application, a linear regression analysis (Kutner et al., 2005) is used to test the association between findings of the automated inductive analysis and students' academic majors.

The automation framework is flexible in terms of which theory can serve as reference and guidance for analyzing qualitative data, i.e., a theoretical framework of a study (Hsieh & Shannon, 2005). This can pave the way for adopting theories that might not have been applied in EE research, especially with the automation of inductive qualitative analysis. For example, the inductive approach allows EE researchers to contextually investigate data and then link findings to theories that explain such phenomena.

1.7 Organization of Dissertation

The rest of this dissertation is organized as follows; the next chapter presents a literature review of EE, the importance of qualitative methods, mind maps as a qualitative data collection tool, QCA and automated text analysis. The automation framework is then introduced in Chapter 3, which includes four steps: the selection of a content analysis approach, mind maps collection and preprocessing, automated analysis, and validation.

Chapter 4 presents the methodology, where the automation framework is tested in two experiments and applied to analyze two different datasets of students' mind maps. Chapter 5 shows preliminary statistics, experiments' findings, and conclusion of findings. In Chapter 6, the discussion, conclusion, and suggestions for future research are presented.

CHAPTER 2

REVIEW OF THE LITERATURE

2.1 Entrepreneurship Education

The first entrepreneurship course started in 1947 at Harvard Business School with 188 graduate students. Ever since, the number of students taking part in entrepreneurship courses has been increasing dramatically. It was estimated that as many as 120,000 American students enrolled in entrepreneurship courses in 2000 (Katz, 2003). EE is now taught at more than 3,000 institutions around the world (Morris & Liguori, 2016). This remarkable growth can be linked to several factors including the realization of policymakers of the significance of entrepreneurship in improving the socio-economic infrastructure of a country (Fayolle, Gailly, & Lassas-Clerc, 2006; Solomon & Matlay, 2008). For example, the European Commission initiated The Entrepreneurship 2020 Action Plan, which states that entrepreneurship is making the European economy more innovative and competitive, allowing for the birth of new companies and enterprises that become the most important source of new jobs and employment (Curth et al., 2015). Further, the European Commission believes entrepreneurship education can boost the European economy to compete globally, stimulate economic growth, and create jobs (Curth et al., 2015). Other countries also recognize the importance of entrepreneurship as drivers for local and global economic development (Acs & Audretsch, 2010).

The appreciation of entrepreneurship education has led to a growth in the supply and demand for entrepreneurship education. Because of the growing investment and expansion of EE, policymakers and educators need to evaluate their return on investment and estimate the impact of EE on stakeholders, i.e., community and students (Souitaris, Zerbinati, & Al-Laham, 2007). Researchers in the field undertake this responsibility and put an effort to measure the impact of entrepreneurship education on students (Farhangmehr, Gonçalves, & Sarmento, 2016; Oosterbeek, Van Praag, & Ijsselstein, 2010; Vanevenhoven & Liguori, 2013; Vesper & Gartner, 1997).

Further, researchers seek to understand the role that some related EE factors, whether internal such as curriculum and method of teaching (Liguori et al., 2018), or external such as academic majors (Kolvereid & Moen, 1997) and cultural norms (Bandera et al., 2018a) can play in forming outcomes. These studies are essential because they guide educators to adjust and improve EE to increase the chances of getting a successful EE impact. For example, findings of these studies can support the design of specialized EE courses and programs among different cultures and academic majors to enhance the effect of factors on the impact of EE (Katz, 2003; Maresch et al., 2016).

2.2 The Importance of Qualitative Research

Research methods in EE are both methodologically and conceptually fragmented. Blenker et al. (2014) find that the methods applied in EE research are grouped into two clusters; quantitative studies of the effect and extent of entrepreneurship education and qualitative single-case studies of different programs and courses. Blenker et al. add that despite the comparability and generalizability of results obtained by quantitative methods, these methods suffer from a limited estimation of measurement variance. On the other hand, qualitative methods are full of contextually sensitive explanations and best pedagogical practices (Blenker et al., 2014). Rahman (2017) also indicates that quantitative research methods, in general, tend to overlook participants' experiences as well as what they mean by something; instead, they only grasp a snapshot of a phenomenon and not in-depth overview (Rahman, 2017).

Although qualitative methods comprise some disadvantages, including being time-consuming and labor-intensive (Patton, 2014), qualitative methods possess strengths for language assessment and testing, including more in-depth insights into interpreting related materials (Rahman, 2017). Researchers using qualitative methods can look directly into the participant's understanding, feelings, perceptions, and behavior (Rahman, 2017). These advantages of qualitative methods are neglected in EE impact studies (Lorz, Mueller, & Volery, 2013). EE research would thus benefit from qualitative methods, not only because there has been insufficient use of them, but because of the advantages these methods offer.

Some critical questions in entrepreneurship can only be asked through qualitative methods (Gartner & Birley, 2002). Mullen, Budeva, and Doney (2009) call for a robust methodological foundation built on state-of-the-art technologies to support further EE paradigmatic expansion and success.

2.3 Mind Map as a Qualitative Data Collection Tool

The substantial use of mind maps in classrooms and related learning and knowledge sharing contexts, e.g., training, meetings, problem-solving discussions, has demonstrated that educators can gain numerous benefits by applying visual mapping techniques that stimulate the diagrammatic construction of knowledge and experience map (Eppler, 2006). For example, researchers have used mind maps to facilitate and evaluate student learning in fields including primary education (Akinoglu & Yasar, 2007), the social sciences (Budd, 2004), engineering (Zampetakis, Tsironis, & Moustakis, 2007), and business (Bandera et al., 2018b).

A researcher uses mind maps as an open-ended survey instrument "*when the researcher does not want to impose bias or suggest relationships by forcing the data into a preconceived coding scheme*" (Jackson & Trochim, 2002, p. 333). Davies (2011) adds that central subjects, where questions are placed in mind maps, might be intentionally ambiguous to permit students to deconstruct the meaning of concerns. Esses and Maio (2002) put forward that such a measure can be used to express positive and negative beliefs and attitudes about the central subject, i.e., attitude ambivalence. They also state that the agnostic nature of mind maps can reduce participants' desire to express what they think the researchers view as favorable.

Mind maps exceed the other qualitative data collection tools from other aspects. One of the most exceptional advantages of using mind maps over conventional qualitative data tools is speed. Krueger (2006) claims that it takes 6 to 8 hours to transcribe an hour and a half focus group, and that may produce 30 or more pages of data to be analyzed line by line (Burgess-Allen & Owen-Smith, 2010). In contrast, Burgess-Allen and Owen-Smith (2010) provide an approach of using mind maps that involves no transcription, and the generation of coded categories happens 'life' during the focus group itself. Therefore, mind maps can be valuable tools for collecting qualitative data (Wheeldon & Ahlberg, 2017).

Recent studies have attempted to reveal the prospective efficiency of mind map topology. Beel, Gipp, and Stiller (2009) suggest that because of their unique topology, mind maps can be used for information retrieval of document summarization, research engine (Beel, Gipp, & Stiller, 2009), and document relatedness (Beel & Gipp, 2010). For qualitative research, mind map topology provides a best-suited tool because it facilitates qualitative analysis by identifying codes, categories, concepts, and connections in the data analysis stage (Wheeldon & Ahlberg, 2012), mainly if a flexible method like qualitative content analysis is used for the analysis (Schreier, 2012).

2.4 Qualitative Content Analysis

The concept of content analysis first emerged in 1952 with the work by Berelson (1959). Content analysis later became a research methodology in mass communication (Prasad, 2008). The early applications of content analysis adopted quantitative methods where inferences were made by quantifying frequencies and recognizable parts of a text, sometimes referred to as manifest content (Hsieh & Shannon, 2005; White & Marsh, 2006). However, objections were raised against quantitative methods because of a superficial analysis that did not take latent contents and contexts into account (Hsieh & Shannon, 2005; Kracauer, 1952). Qualitative approaches to content analysis were subsequently developed (Altheide & Schneider, 1996; Hickey & Kipping, 1996; Mostyn, 1985), and content analysis has become a dominant qualitative research technique in empirical social sciences including communications, psychology, political science, history, and language studies (Prasad, 2008).

QCA is one of several research methods used to analyze text data (Mayring, 2004). Other methods include grounded theory (Corbin & Strauss, 2014), ethnography (Aktinson & Hammersley, 1998), and phenomenology (Giorgi, 2009). The focus of QCA is on language characteristics as a communication tool with consideration of the text's contextual meaning (Lindkvist, 1981; McTavish & Pirro, 1990; Tesch, 2013). It is defined as "*a* *research method for subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns*" (Hsieh & Shannon, 2005, p. 3). Researchers regard QCA as a flexible method for analyzing text data in a systematic way that assigns pieces of the investigated material into categories of a coding scheme (Schreier, 2012). Weber (1990) states that QCA can make valid inferences from texts and thus can be used to code open-ended survey questions. The following examples present studies from different disciplines that used QCA as a qualitative method (Mayring, 2004):

- Vicini (1993) has conducted 14 open-ended in-depth interviews with educational advisors about detailed case studies from their advisory service to reconstruct their theory of mind of advice.
- Beck and Vowe (1995) have analyzed 25 media products (newspapers, journals, radio transmissions) concerning new multimedia approaches.
- Bauer et al. (1998) analyzed 21 Alzheimer's disease patients' biographies to find common patterns and compare them with 11 vascular dement patients of the same age.

Hsieh and Shannon (2005) introduced three approaches for QCA: directed, conventional, and summative. All three interpret meaning from text data. The significant differences among them are the source of codes, codding schemes, and trustworthiness. The directed and conventional approaches are extensions of the traditional QCA approaches, where the former refers to the deductive approach, and the latter refers to the inductive approach (Hsieh & Shannon, 2005). Mayring (2014) argued that although the number of procedures of QCA has been developed, the two main approaches are still deductive and inductive analysis. For these reasons, this dissertation focuses on automating the deductive and inductive QCA of mind maps. The two approaches are explained in detail in Chapter 3.

2.4.1 Automated Qualitative Content Analysis

Computer software evolution for text data processing has enabled computational and automated content analysis to flourish (Krippendorff, 2018). One of the first applications of computers to content analysis was the system developed by Stone and Hunt (1963). Developments in other fields, including psychology and linguistics research, also advanced the use of computers in content analysis (Schank & Abelson, 1977). In the last two decades, a substantial contribution to the automation of content analysis comes from the interdisciplinary fields of statistical Natural Language Processing (NLP) (Manning & Schütze, 1999) and Machine Learning (Bishop, 2006).

Manning and Schütze (1999, p. 31) define statistical NLP as "*all quantitative approaches to automated language processing, including probabilistic modeling, information theory, and linear algebra.*" Alpaydin (2009, p. 3) defines machine learning more simply as "*programming computers to optimize a performance criterion using example data or past experience.*"

Researchers in political science (Grimmer & Stewart, 2013; Hillard, Purpura, & Wilkerson, 2007), sociology (Shor et al., 2015), psychology (Tausczik & Pennebaker, 2010), and other social sciences (Carley & Roberts, 1997) have used machine learning techniques to automate content analysis to measure constructs in a systematic way that humans may not detect. Thus, in this research, state-of-the-art machine learning models are used to automate the qualitative content analyses of mind maps used in EE research.

CHAPTER 3

THE AUTOMATION FRAMEWORK

3.1 Introduction

The automation framework is novel and intended to automate the content analysis of mind maps used as an open-ended survey tool in EE settings and research. It enables EE researchers to employ mind maps as a data collection tool, select a QCA approach that suits their research aims, and automate the data analysis process. The automation framework can also be utilized in other academic disciplines, such as communication studies, public health, and psychology. The framework includes four sequential steps: selecting a QCA approach, collecting mind maps and preprocessing, automated analysis, and validation, reliability, and model evaluation. Figure 3.1 shows the automation framework diagram.

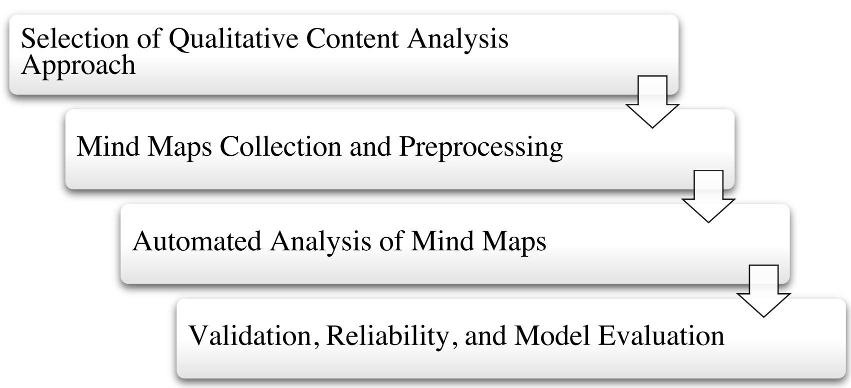

Figure 3.1 The automation framework diagram including the four steps.

3.2 Selection of Qualitative Content Analysis Approach

The selection of a specific approach depends on the problem being investigated by researchers (Weber, 1990). The two approaches of qualitative content analysis, deductive and inductive, are traditionally applied in different settings. For example, whereas the inductive approach does not require prior knowledge to begin the analysis, the deductive approach requires a theory or relevant research findings, e.g., a reference model.

Krippendorff (2018) states that content analyses may start with data that are not meant to answer specific research questions. Instead, data can be read, understood, and interpreted to form meaningful structures and units. Thus, if necessary, selecting a qualitative analysis approach can be done after collecting mind maps.

The two QCA approaches are presented next along with descriptions of how to conduct them manually for mind maps analysis.

3.2.1 The Deductive Approach

The deductive approach to content analysis brings prior formulated knowledge and theoretical aspects to connect with the text (Mayring, 2004). It can be applied to extend or confirm a theory (Hsieh & Shannon, 2005). It requires research findings or a theory to serve as guidance; this study refers to that as a reference model. An extant theory also helps develop research questions and provides definitions for variables and the relationships among them, which then determines the initial scheme of coding (Hsieh & Shannon, 2005). The deductive approach holds a more organized process than the inductive approach (Hickey & Kipping, 1996), because the identification of critical categories is carried through insights of the reference model (Potter & Levine-Donnerstein, 1999).

The deductive approach's goal is to provide explicit coding rules, definitions, and examples for each category. The definition of categories are included within a coding plan (Mayring, 2004). Category definitions, prototypical text passages, and rules for distinguishing different categories are formulated based on the reference model, completed step by step, and revised with the analysis (Krippendorff, 2018; Mayring, 2004).

One of the main strengths of the deductive approach is that it can support and extend an existing theory. It is also less complicated than the inductive approach in that it provides direct definitions of coding schemes and categories. A risk in using the deductive approach is that researchers may define passages of text that are only supported by the reference model and neglect important contextual aspects of the data (Hsieh & Shannon, 2005). Mind maps mitigate this risk by giving a researcher already divided passages of text, i.e., nodes inside mind maps (Tattersall, Watts, & Vernon, 2007; Wheeldon & Ahlberg, 2017). Thus, no parts of the data are missed since analysts can code all nodes.

3.2.1.1 Manual Deductive Analysis of Mind Maps. Analysts can choose one of two strategies to begin manual coding under the deductive approach (Krippendorff, 2018). In the first strategy, an analyst highlights words of interest in the mind maps and codes these marked words using the reference model. This strategy is useful when the goal of the study is to find and categorize all observations of a specific phenomenon (Krippendorff, 2018). For mind maps, this means analysts can code nodes that only reflect the studied cases and leave the remaining nodes unclassified.

The second strategy requires coding the entire text into the pre-defined classes. Hard to code texts are analyzed to check if the existing reference model can provide a subcategory. This approach can be chosen when an analyst is convinced that all observations and categories are equally important, and an immediate coding does not deviate the analysis (Krippendorff, 2018). For mind maps, this strategy leads to a complete classification of all nodes.

The use of mind map as a data collection tool facilitates manual deductive because nodes inside a mind map typically represent unique and independent thoughts, experiences, and ideas linked to the central concern (Wheeldon & Ahlberg, 2017). Hence, nodes by themselves are separate and individual segments of the text that indicate specific categories (Tattersall, Watts, & Vernon, 2007). They are respondent-made categories, which hold a firmer foundation when compared to categories made by an analyst. Mayring (2014) points out that there is a coding unit in content analysis, which is essential for qualitative analysis because it is the smallest part of texts that can be coded into one category, i.e., documents in the context of text analytics (Aggarwal & Zhai, 2012). Mind maps provide a direct coding unit for qualitative analysts, namely nodes.

3.2.2 The Inductive Approach

The inductive approach is appropriate when prior knowledge relevant to the phenomenon under study is fragmented or limited (Elo & Kyngäs, 2008). In the inductive content analysis, researchers extract codes, categories, or themes directly from the text data without a reference model (Hsieh & Shannon, 2005). Unlike the deductive approach, predefined categories are avoided and instead are defined by the data (Kondracki, Wellman, & Amundson, 2002). It is often applied as the first step in qualitative methods, such as grounded theory (Krippendorff, 2018).

The traditional quantitative content analysis holds few answers to the question of where the categories come from, and how the generative process of categories is developed. However, this is a focal interest within the framework of qualitative content analysis. It develops a systematic way of extracting categories and carrying on with the perspective of interpretation as close as possible to the investigated materials; "How categories are defined ... is an art. Little is written about it" (Krippendorff, 1980, p. 76). QCA includes established procedures of inductive category development related to the reductive processes mapped out within the psychology of text processing (Mayring, 2004).

The goal of the inductive process is to design a criterion of definition acquired from research questions to map out the facets of the textual content under study (Krippendorff, 2018). Following this benchmark, the content is analyzed through, and categories are step by step formulated. Those categories are then revised with feedback cycles to arrive at reliable categories eventually (Mayring, 2004).

A critical advantage of the inductive approach is that it gains information directly from study samples without imposing predetermined categories or theoretical standpoints (Krippendorff, 2018). On the other hand, a principal challenge of this approach is that results may not present adequate assimilation of the context, hence failing to find meaningful categories (Krippendorff, 2018). This framework provides validity and reliability guidelines for automating the inductive analysis of mind maps.

3.2.2.1 Manual Inductive Analysis of Mind Maps. A manual inductive content analysis of mind maps starts with highlighting keywords to create an initial impression (Mayring, 2004). That initial step is followed by labeling keywords to create a coding scheme, where categories are established. Categories are used to organize the data into clusters; Krippendorff (2018) recommends using graphs and diagrams to visualize clusters and categories. Structure of mind maps offers significant value for analysts by that means because a mind map is a diagram by itself. Instead of looking at regular passages of texts to search for categories within the data, mind maps allow for direct identification of categories and their connections based on participants' inputs (Tattersall, Watts, & Vernon, 2007; Wheeldon & Ahlberg, 2017).

Finally, a definition of each emerged category or cluster is developed, then analysts can measure and find relationships among groups of interest (Krippendorff, 2018).

3.3 Mind Maps Collection and Preprocessing

The use of mind maps presents a participant-led method to aid both the deductive and inductive analysis process (Serry & Liamputtong, 2013; Tattersall, Watts, & Vernon, 2007; Wheeldon & Ahlberg, 2017). Utilizing mind maps to collect individualistic and personalized data from studies participants is a typical use of mind maps in research (Tattersall, Watts, & Vernon, 2007; Wheeldon & Ahlberg, 2017).

A mind map in this framework is defined as a single theme map used to answer one question. For another question, another mind map is generated. Students can generate mind maps online; there are free and available to download mind map applications. Students should learn beforehand how to use mind maps to answer an open-ended question placed in the center of mind map (Bandera et al., 2018b).

Once mind maps are collected, they are preprocessed. Data preprocessing is essential in the automated text analysis process (Manning & Schütze, 1999). Data preprocessing involves uploading mind maps into the software, converting raw data to machine-readable file formats if necessary, cleaning, normalizing, stemming, and prepare the data for the automation process.

Data preprocessing includes normalizing texts, e.g., tokenizing, which means segmenting text into linguistic units such as words, numbers, and punctuations. The smallest textual unit in this framework is called token; this is different from the unit of analysis, i.e., documents, which is the block of texts that can be coded (Mayring, 2014).

Tokenization allows for further text preprocessing, such as stemming. The primary goal of stemming is to reduce various grammatical forms of a word like its noun, verb, adjective, and adverb to its root (Jivani, 2011). For instance, instead of having love, loving, loved as different words and then features, stemming combine them under the term love. Stemming has been customary in text analytics (Aggarwal & Zhai, 2012). The Porter stemmer for English has been the most popular algorithm (Jivani, 2011). Stemming assists the automated text analysis because it decreases the sparsity of generated matrices for texts. Chapter 4 presents further preprocessing functions conducted in this study, including how data is transferred into textual feature representations.

3.4 Automated Analysis of Mind Maps

In this step, researchers apply machine learning algorithms to automate QCA of collected and preprocessed mind maps. Figure 1.2 shows the diagram of the automation models corresponding to the two QCA approaches. Text classification models automate the deductive approach of content analysis, and STM automates the inductive approach to content analysis.

3.4.1 Automated Analysis of Deductive Approach

3.4.1.1 Text Classification. Text classification is the automated model of textual or partially textual entities that include information retrieval, categorization, and filtering, in addition to, Natural Language Processing tasks such as word tagging (Lewis & Gale, 1994; B. Pang & Lee, 2008). Text classification is the task of assigning a value from predefined classes, whether a binary or multi-class, see Table 3.1, to each document $\{d_j\} \in D$, where D represents a domain of documents. Data in text classification includes a pair of $\{d_j, c_i\}$ where $C = \{c_1, \dots, c_i\}$ is a set of predefined classes. A training set of n manually labeled documents $\{d_1, c_1\}, \dots, \{d_n, c_n\}$ is provided to the classification model. More formally, we seek to approximate the undefined target function $\{D\} \rightarrow \{C\}$, which is called the classifier and where C here is unknown and therefore predicted (Sebastiani, 2002).

Many algorithms used in text classification includes Naïve Bayes, Logistic Regression, Decision Trees, Support Vector Machine, and K-nearest Neighbor. However, selecting which algorithm to use is task dependent, as many studies have observed, no single classification algorithm is the best for all cases and datasets (Salzberg, 1997). Hartmann et al. (2019) recommend to test available algorithms to recognize the best performance (Hartmann et al., 2019). Worth noting is the limited amount of textual data in mind maps, which is generally challenging in text classification (Sebastiani, 2002).

Table 3.1 Types of Classes in Text Classification

Class Type	*Description*
Binary	The document is to be classified into one, and only one, of two non-overlapping classes under one output (Y_1:C_1 *or* C_2), e.g., spam or not spam under type of mail.
Multi-class	The document is to be classified into one, and only one, of more than two non-overlapping classes under one output Y (Y_1: $C_1, \dots, C_i$), e.g., sport, politics, or business under category.
Multi-labelled	The document is to be classified into several classes of two or more outcomes ({Y_1: $C_1, \dots, C_i$} ; {Y_j: $, \dots, C_n$}) e.g., politics and international under category and region.
Hierarchical	The document is to be classified into one, and only one, class which are grouped into super classes or divided into subclasses, e.g., sports > soccer.

3.4.1.2 Feature Representation. The main concept in feature representations is to find the best way to represent text as quantitative data, i.e., real numbers, in order for machine learning models to achieve better results (Goldberg). The Bag-of-Words (BOW) model is one of the most popular textual representations in classification tasks (Zhang, Jin, & Zhou, 2010), in which terms in the corpus, i.e., collection of all documents, are used as features to represent document content. It produces a document term matrix, where rows and columns represent documents and terms, respectively.

A variety of methods can be used to compute the weight or value of features in the BOW model matrix. A simple one is a binary method, where 1 and 0 stand for the term occurrence and no occurrence. Weights can also be computed by applying Term-Frequency Inverse Document-Frequency (TF-IDF). The TF part computes the number of times that term i occurs in document j (Wu et al., 2008). IDF calculates the rareness of a term through all documents and most frequent terms get less weight. It is computed as follows:

$$For\ a\ term\ i\ in\ document\ j:$$

$$w_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right) \tag{3.1}$$

$tf_{i,j}$ = number of occurrences of i in j

df_i = number of documents containing i

N = total number of documents

TF-IDF method does so to ensure that most frequent terms do not dominate while supplying little information (Manning & Schütze, 1999).

Bloehdorn and Hotho (2004) argue that TF-IDF is limited to detecting patterns in the used terminology only, while contextual patterns are still not detected. Therefore, Normalized Pointwise Mutual Information (NPMI) can be applied instead of the TF-IDF algorithm to grasp conceptual meaning among words (Bouma, 2009). It computes word to word similarity by evaluating which words co-occur in the same context, where an entire corpus represents a context. It is computed as follows:

$$NPMI(v,t) = -log\ P(v,t) \cdot log \frac{P(v,t)}{P(v).P(t)} \tag{3.2}$$

- v and t are two words in the corpus.

From this, we obtain an embedding of words where values are not only assigned by a mere occurrence but also by co-occurrence with other words. Because it is normalized, the values range between -1 and 1, where -1 means never occurring together and 1 for total occurrence (Bouma, 2009).

Other advanced textual representations, such as neural network word-embedding and Word2Vec (Mikolov et al., 2013), are used to feed innovative machine learning models, such as deep learning. They achieve impressive results for classification tasks; however, deep learning models perform well with a vast quantity of data (Najafabadi et al., 2015).

In this work, three weighting methods, TF, TF-IDF, and NPMI, for the BOW model, are tested.

3.4.1.3 Feature Extraction. Feature extraction in text classification means the generation of new features from the existing ones (Blum & Langley, 1997). Feature extraction can include feature selection, dimensionality reduction, and feature engineering (Guyon et al., 2008).

Feature extraction techniques offer practical benefits to text classification tasks when implemented as feature reduction technique; they reduce sparsity in the data, save memory storage, and shorten computation time (Humphreys & Wang, 2017). Standard feature reduction techniques include Principal Component Analysis (Wold, Esbensen, & Geladi, 1987), word clustering (Slonim & Tishby, 2000), and Singular Value Decomposition (SVD) (Golub & Reinsch, 1971).

The generation of new features based on link analysis among nodes of a mind map can assist classification models to reach better performance (Humphreys & Wang, 2017). The automation framework implements two different methods for generating features. The first method emphasizes link analysis among nodes of mind maps based on their branch, hierarchal level, and the number of nodes at the same level. The second method is focused on link analysis among nodes of mind maps based on their specific map and branch. This research includes hypothesis testing for each method.

3.4.2 Automated Analysis of Inductive Approach

Unsupervised learning models discover patterns from data without labeling outcomes (Friedman, Hastie, & Tibshirani, 2001; Hinton, Sejnowski, & Poggio, 1999). The inductive approach to content analysis also discovers categories directly from data without using prior knowledge (Hsieh & Shannon, 2005).

Topic modeling, a common unsupervised learning approach, is a statistical-based algorithm used to identify latent topics within a corpus of text (Blei, Ng, & Jordan, 2003; Wallach, 2006). Topic modeling is appropriate when a researcher wants to inspect the text without prior categories of words. It is applied to recognize words that tend to occur together through documents. A group of words is referred to as a "topic" (Wallach, 2006). It is useful for researchers when a reference model of texts does not exist or is hard to apply (Humphreys & Wang, 2017). Topic modeling is especially helpful in cases where analyzing and classifying texts, or even a subset of the texts is costly, complex, restricted (Humphreys & Wang, 2017). For example, Mankad et al. (2016) use topic modeling to analyze hotel reviews, and they discover that the reviews mainly mention five topics, which are "amenities," "location," "transactions," "value," and "experience." When topics are identified, a study of their relationships with each other can be done with other variables (Roberts et al., 2014).

Algorithms of topic modeling include Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003), Hierarchal Dirichlet Process (Teh et al., 2005), and Structured Topic Model (STM) (Roberts et al., 2014).

3.4.2.1 Latent Dirichlet Allocation. LDA is one of the traditional topic models. Blei (2012), the author of LDA, explains the assumptions of the model as follows:

- *Topics* are a fixed number of groups of *terms* that tend to occur together in *documents*
- Each *document* in the corpus is made from these *topics*

LDA uses the following joint distribution of the hidden and observed variables:

$$p(\beta_{1:k}, \theta_{1:d}, z_{1:d}, w_{1:d})$$
$$= \prod_{i=1}^{k} p\,(\beta_i) \prod_{d=1}^{D} p\,(\theta_d) \tag{3.3}$$
$$\left(\prod\nolimits_{n=1}^{N} p\,(z_{d,n} |\, \theta_d) p(w_{d,n} |\, \beta_{1:k}, z_{d,n})\right)$$

where the topics are *k*, and each β_k is a distribution of vocabularies over topics. The topics probability for the *d*th document are denoted as θ_d, where $\theta_{d,k}$ is the probability for topic *k* in document *d*. The word-topic assignment for the *d*th document are z_d, where $z_{d,n}$ is the assignment for the *n*th word in document *d*. Finally, the observed words for document *d* are w_d, where $w_{d,n}$ is the *n*th word in document *d*, which is an element from the fixed vocabulary (Blei, 2012).

Note that this distribution specifies a few dependencies. $z_{d,n}$ is depending on the topic probability per document θ_d. Also, the observed word $w_{d,n}$ is conditional on both the topic distribution over vocabularies $\beta_{1:k}$ and the word-topic assignment $z_{d,n}$ (Blei, Ng, & Jordan, 2003). To estimate these probabilities based on data and number of topics, different methods including Variational Inference and Gibs Sampling are used (Blei, Ng, & Jordan, 2003).

Figure 3.2 shows how documents, mind map nodes in this research, can be defined as a distribution over topics, and mind maps can be defined as a distribution over nodes.

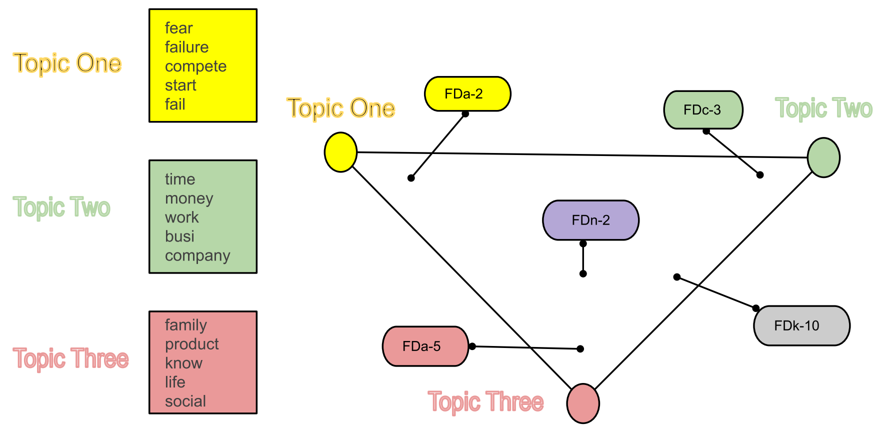

Figure 3.2 The assumptions of LDA for mind maps. Each node in a mind map is distributed over topics that are distributed over words. The simplex depicting nodes distribution over topics as probability, e.g., [.3, .5, .2]. The line from each node's ID shows its position in the topic space.

3.4.2.2 Structural Topic Model. STM is a mixed-membership topic model, like LDA. The goal of STM is to discover topics and estimate their association with document-level metadata (covariates), such as affiliation or treatment (Lucas et al., 2015; Roberts et al., 2013). For qualitative studies, STM enables observation of treatment effects or variables of interests and hypothesis testing (Roberts, Stewart, & Tingley, 2014). For instance, if two distinct groups that draw mind maps, the STM permits direct observation and comparison between the two groups' latent topics. The authors of STM states that the model "*can fruitfully be used at either an exploratory stage prior to using human coders or as part of making credible inferences about the effect of treatments/frames/covariates on the content of open-ended responses"* (Roberts et al., 2014, p. 3). Mind maps in this framework are utilized as an open-ended survey instrument.

Structural Topic Model has a few advantages over LDA that make it a popular choice for social science research, particularly for open-ended survey analysis (Roberts et al., 2014). In addition to estimating a topic prevalence measure, which is the proportion of words contributable to each topic, STM computes the most likely words generated for each covariate, providing a topical content measure (Roberts et al., 2014). To explain further, in LDA, the collection of documents is assumed to be unstructured, meaning that each document is generated from the exact generating process regardless of any additional information that each document possesses, such as the author's affiliation (Blei, 2012). On the other hand, STM is built to incorporate any additional information about the document into the evaluation process; this extra feature of STM permits estimating of methodical changes in both topical prevalence and topical content (Roberts et al., 2014).

Another major factor for using STM to automate inductive content analysis of mind maps is the text size. Mind maps consist of short texts; for example, Joeran Beel and Langer (2011) analyze about 19 thousand mind maps and find that maps have an average of 30 nodes per map and 4.8 words per node. In recent years, researchers have examined the use of topic modeling on the short text (Cheng et al., 2014; Quan et al., 2015; Zuo, Zhao, & Xu, 2016) and conclude that specific topic models can perform better than the traditional models when handling short text. Lucas et al. (2015) and Reich et al. (2014) use STM to analyze Twitter feeds and online class forum and report that STM finds syntactic patterns with semantic meaning, identify differences in patterns across users, and discover most unique terms that exemplify the covariance within a topical pattern. The two studies' data contain a similar text size to that in mind maps.

Structural Topic Model also has a technical advantage over other topic modeling algorithms when estimating the posterior probability, which is intractable and problematic in all mixed-membership topic models (Lucas et al., 2015). STM attempts to solve for that by changing its initialization. One of the two used approaches by STM is the method of moments, which is globally consistent and deterministic under fair conditions (Arora et al., 2013; Roberts, Stewart, & Tingley, 2016). It is known as a spectral initialization because it implements a Non-negative Matrix Factorization model as an initialization. STM authors find this to be very helpful and produce better results consistently (Roberts, Stewart, & Tingley, 2014).

For each document d with vocabulary of size V, an STM model with K topics is summarized as follows:

1. Draw the document-level attention to each topic from a logistic-normal generalized linear model based on a vector of document covariates X_d

$$\sim \vec{\theta}_d | X_d\gamma, \Sigma \approx \ LogisticNormal(\mu \ = \ X_d\gamma, \Sigma) \tag{3.4}$$

   where X_d is a p-by-1 vector, γ is a p-by-K −1 matrix of coefficients and Σ is K −1-by-$K - 1$ covariance matrix.

2. Form the document-specific distribution over words representing each topic k using the baseline word distribution m, the topic specific deviation κ_k, the covariate group deviation κ_g and the interaction between the two κ_i.

$$\beta_{d,k} \ \alpha \ exp(m \ + \ \kappa_k \ + \ \kappa_{g_{\,d}} \ + \ \kappa_i = (_{k,g_{\,d}})) \tag{3.5}$$

   m, and each κ_k, κ_g and κ_i are V-length vectors containing one entry per word in the vocabulary.

3. For each word in the document, $(n \ \in \ 1, \ldots, N_d)$:

- Draw word's topic assignment based on the document-specific distribution over topics.

$$z_{d,n} | \vec{\theta}_d \; Multinomial(\vec{\theta}_d) \tag{3.6}$$

- Conditional on the topic chosen, draw an observed word from that topic.

$$w_{d,n} | z_{d,n}, \beta_{d,k} =_{z_{d,n}} \; Multinomial(\beta_{d,k} =_{z_{d,n}}) \tag{3.7}$$

In STM, a partially collapsed variational Expectation-Maximization algorithm gives estimates of the model parameters upon convergence (Roberts, Stewart, & Tingley, 2014). For γ, κ, and optionally Σ, regularizing prior distributions are used to improve interpretation and stop overfitting.

Each topic generated by STM includes four distinct types of words. They are:

- Highest Prob: the highest probability words inside each topic, which are inferred directly from topic-word distribution parameter β (Roberts, Stewart, & Tingley, 2014).
- FREX: the frequent and exclusive words inside each topic. These words differentiate topics "exclusive" (Roberts, Stewart, & Tingley, 2014).
- Score: a score for words defined by:

$$\beta_{w,k} \left(\log \beta_{w,k} \; - \frac{1}{K \sum k}, \log \beta_{w,k'} \right) \tag{3.8}$$

(Chang & Chang, 2010)

- Left: is calculated by dividing the topic-word distribution denoted as β_k by the empirical word count probability distribution (*wbar)*:

$$Left = \beta_k / wbar \tag{3.9}$$

(Taddy, 2012).

Thus, STM, offer analysts several options to choose the type of words summarization for latent topics.

3.5 Validation, Reliability, and Model Evaluation

Validity deals with the truthfulness of analysis and its findings, while reliability refers to the degree of which an analysis led to the findings are reproducible and consistent (Altheide & Johnson, 1994; Crook et al., 2010; Krippendorff, 2004). At this step of the automation framework, the validation and reliability of the automated analysis are evaluated.

The inspection of validation and reliability in this framework follows the four properties presented by Egami et al. (2018): inclusive, reflexive, and applicable to the automated deductive and inductive qualitative content analyses of mind maps. The four are interpretability, theoretical interest, fidelity (validity), and tractability (reliability). Within the scope of each property, explanations and examples of the automated analysis of mind maps are discussed.

First, interpretability, as research and data specific, means that readers should recognize what an analysis is measuring. In the context of the automation of deductive analysis performed by text classification, this property entails that the predefined classes used to label mind maps are understandable when linked to their theory, i.e., a reference model. For the automated inductive analysis performed by STM, this assures that the generated STM latent topics are interpretable, i.e., the semantical meaning of topics is explicit (Egami et al., 2018). In the application of STM introduced in this work, most representative documents and FREX words of latent topics are used to guide topics' interpretation.

Second, the automated qualitative analysis should include measures that are not only interpretable but also demonstrative of theoretical interest (Egami et al., 2018). In text classification, the theoretical interest is established when the analysis involves prior knowledge or a theory, i.e., a reference model. Thus, the theoretical interest of text classification is fundamentally derived from the reference model. In the automation of inductive analysis, this concern questions whether the STM latent topics represent a theoretical interest or not (Roberts et al., 2014). The theoretical interest of STM latent topics is reflexive to the third property "fidelity," because defining STM latent topics is similar to developing manual inductive categories, where both are theory-driven, leading to the desired theoretical interest (Mayring, 2014; Potter & Levine-Donnerstein, 1999).

Third, fidelity means validity in measurement of content analysis (Grimmer & Stewart, 2013; Krippendorff, 2004; Patton, 2014). In content analysis, validity is the extent to which a measuring process conveys the intended and essential aspects of the concept (White & Marsh, 2006). The concern raised by validity in the deductive analysis is whether mind map nodes, i.e., documents, hold a content that is implied by assigned labels; it questions the truthfulness of assigning documents to labels (Krippendorff, 2018). Besides validating classification models, which is described in the next subsection, validity is pertinent to the manual coding of the training set, in which assignment of labels to mind maps must be indicative of their content.

Structural Topic Model validity is related to how the STM topic proportion per document represents its content and evaluating the analyst's definition of discovered STM latent topics. STM fidelity deals with the appropriateness of choosing the number of topics, i.e., *k*, which is a critical choice in unsupervised learning models (Roberts, Stewart, & Tingley, 2014).

Because *k* should be determined before conducting the analysis, STM offers two metrics to evaluate the quality of captured topics for any given *k* (Grimmer & Stewart, 2013; Roberts, Stewart, & Tingley, 2014). The two metrics are semantic coherence and inclusivity. Semantic coherence is developed by Mimno et al. (2011), which is related to pointwise mutual information (Lau et al., 2010). A higher score means that the most probable words in a particular topic frequently co-occur, reflecting better semantic meaning to topics. Mimno et al. (2011) present that semantic coherence corresponds well with a human measure of topic quality.

Exclusivity refers to the weighted mean of the words in terms of exclusivity and frequency (Roberts et al., 2014). Topics with exclusive and high frequent words attain a higher score, then the weighted mean is computed for all topics to evaluate the model exclusivity score. More details about their implementation are presented in experiment two in Chapter 4, where the metrics assist in deciding the number of topics, i.e., *k*, inserted into STM.

Finally, the development and deployment of the analysis should be tractable and reproducible. In content analysis, reproducibility is the most important meaning of reliability (Krippendorff, 2018). One of the methods to deal with manual coding's reliability is by having more than one coder to independently code the text and compute the intercoder reliabilities (Prasad, 2008). Bahr, Albrecht, and Chadwick (1984) introduced a formula to calculate intercoder reliability and called it a coefficient of reliability. In the context of this framework, the coefficient of reliability is formulated as follows:

$$Cofficient\ of\ Reliability = \frac{Number\ of\ Coded\ Nodes\ Agreed\ on}{Total\ Number\ of\ Coded\ Nodes} \tag{3.10}$$

the numerator holds the number of coded nodes that both coders assign to the same label, i.e., concurrence assignments. A coefficient value between .81 and 1 indicates the "almost perfect" agreement level (Landis, Koch, 1977).

In the inductive analysis, the reliability of the automated analysis is assured due to STM consistency. As explained earlier, STM uses the method of moments as its initialization, which is consistent and deterministic (Arora et al., 2013; Roberts, Stewart, & Tingley, 2016). However, defining and interpreting the STM latent topics is performed manually, and reliability can be an issue. The STM package in the statistical programming language R includes a function that finds most representative documents for each latent topic, which can be used as guidance for defining topics. This assist coders to enhance reliability (Roberts, Stewart, & Tingley, 2014). STM also generates four different types of words summary for latent topics. Analysts can check these groups to make a consistent interpretation of topics (Roberts, Stewart, & Tingley, 2014).

Table 3.2 shows the validation and reliability properties presented in this section with examples from text classification and STM analyses (Egami et al., 2018).

Table 3.2 The Four Properties of Validation and Reliability of the Automation Framework

EVALUATION	AUTOMATED ANALYSIS	EXAMPLES
INTERPRETABLE	Classification	Predefined classes are interpretable
	STM	Latent Topics are semantically meaningful
THEORETICAL INTEREST	Classification	Classes can be used to test a hypothesis
	STM	Latent topics can be used to test a hypothesis
FIDELITY (VALIDATION)	Classification	Accuracy of a model is computed
	STM	k corresponds to the number of topics discussed in the data
TRACTABILITY (RELIABILITY)	Classification	Manual coding for training set is reliable
	STM	STM results are tractable and reliable

3.5.1 Model Evaluation

In text classification, models can be evaluated by comparing model labels with manual labels (Aggarwal & Zhai, 2012). For models with a binary class, the Receiver Operating Characteristic (ROC) curve has been widely used to measure text classification performance (Bradley, 1997) . The ROC curve is computed as follows:

$$AUC = \frac{1}{2}\left(\frac{tp}{tp + fn} + \frac{tn}{tn + fp}\right) \tag{3.11}$$

where tp and tn stand for true positive and true negative, and fn and fp stand for false negative and false positive, respectively. tp and tn are the true predicted outputs, whether the class is positive or negative, e.g., yes or no. fn and fp are false predicted outputs whether they belong to positive class, e.g., yes, or negative one, e.g., no (Bradley, 1997).

For multi-classification models, accuracy and F-score metrics can be used. Classification accuracy is widely used as a metric computed from the confusion matrix using the testing set to report the significance of differences between true and false predictions (Diebold & Mariano, 2002; Labatut & Cherifi, 2012). Accuracy can be computed as follows:

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn} \tag{3.12}$$

The accuracy measure can be biased when target classes are imbalanced, e.g., there is one dominant class (Davis & Goadrich, 2006).

For imbalanced classification problems, other metrics are suggested in the literature (Sokolova & Lapalme, 2009). One of the recently used metrics is F-score, and it is computed as follows:

$$F - score = 2\frac{precision\,.recall}{precision + recall} \tag{3.13}$$

where precision is computed by dividing tp over $tp + fp$, and recall is computed by dividing tp over $tp + fn$ (Sokolova & Lapalme, 2009).

Depending on the target class type, e.g., binary or multi-classes, the above metrics can be used to compare text classification algorithms implemented to automate the deductive analysis of mind maps.

CHAPTER 4

METHODOLOGY

4.1 Introduction

The automation framework is tested in two different experiments. The first experiment tests the hypotheses of applying text classification to automate deductive analysis of mind maps and the exploitation of mind map topology in such a process. The second experiment examines the hypotheses of using STM to automate inductive analysis of mind maps, comparing STM to LDA, and treating nodes of mind maps as the unit of analysis. Each experiment includes an EE application with case-specific problems. The entrepreneurship applications serve as a method of testing the automation framework's applicability and feasibility in EE research.

The automation framework's steps, presented in Chapter 3, were implemented as the preliminary steps in each case. The first experiment includes subsections of the following: experiment's aim, reference model, participants' definition, data preprocessing of mind maps, training set, automated analysis, measures of hypotheses, and entrepreneurship application. The second experiment includes the following subsections: experiment's aim, participants' definition, data preprocessing of mind maps, automated analysis, measures of hypotheses, and entrepreneurship application.

Before starting with the two experiments and for the sake of avoiding redundancy, a general overview of the mind map collection, initial mind maps preprocessing, and quantitative statistics of mind maps are introduced first. Figure 4.1 shows the flowchart of the methodology chapter.

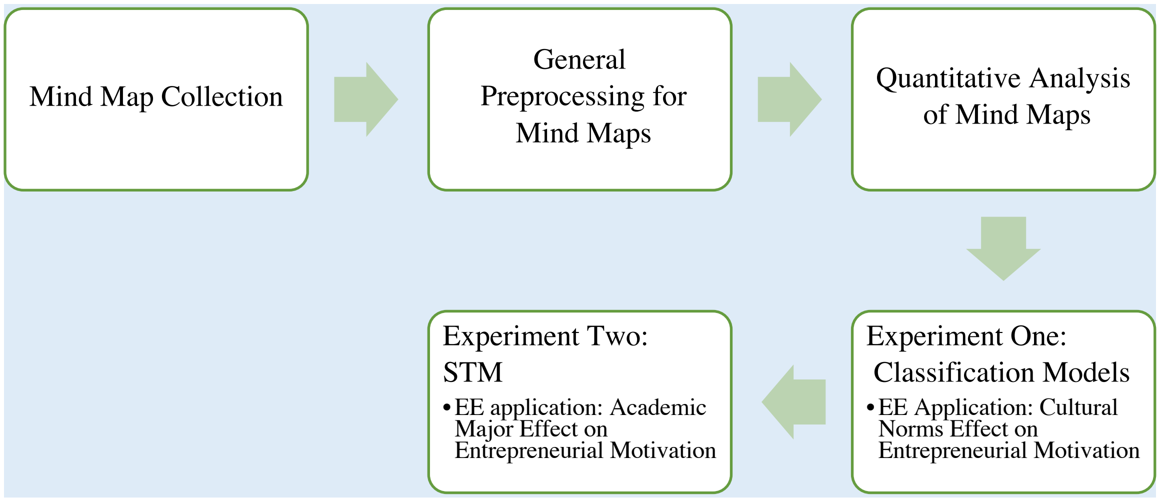

Figure 4.1 A flowchart of how the methodology chapter is presented.

4.2 Mind Map Collection

The first dataset, referred to "culture" for brevity, was collected from students of two cultures, U.S. and France. The second dataset, referred to as "major" for brevity, was collected from students of two colleges, business (BUS) and computer science (CS), over five semesters from a U.S. school.

Although the two datasets were collected in different settings, both were collected using mind maps as a data collection tool and following the same sampling technique. Students enrolled in an entrepreneurship course were asked to draw a mind map answering for what motivates her to create a new venture, and to draw another mind map to answer for what deter her from doing so. Thus, each dataset includes two subsets: referred to as "motivation" and "deterrence" for brevity.

4.3 General Mind Maps Preprocessing

In this dissertation, a cohesive procedure to store the collected mind maps using a plain text file was established because mind maps have already been collected before building the automation framework. The initial preprocessing of mind maps is identical for the two datasets, culture and major. This procedure can be applied whenever the collected mind maps are not in machine-readable formats, such as Portable Document Format (PDF). Students in the samples have submitted their mind maps in different formats, including images and PDF.

A plain-text file was manually created for each collected mind map. Each mind map, whether submitted as an image or a PDF file, was manually written into one text file. For each mind map, there is one plain-text file, where each node was inserted into one. A consistent conversion of mind maps into text files was maintained. Figure 4.2 shows a mind map and explains the sequential order of moving nodes into a text file.

Figure 4.2 Illustration of how a mind map was converted into a text file starting from the top right parent node and moving in a clockwise direction. Numbers in green boxes show the sequential order of writing nodes into lines. These numbers were written only for illustration purposes, and they are not included in the conversion.

This research hypothesizes that the topology of mind maps can be helpful for the automated analysis; thus, the node's level was inserted at the end of their node's line. Levels are represented by numbers starting from 1, and level 1 refers to parents, level 2 for children, level 3 for grandchildren, level 4 for grand grandchildren. Figure 4.3 shows the level of nodes for the mind map in Figure 4.2. These numbers were used to extract features of mind map topology. These features were only used in automating the deductive analysis, i.e., text classification models. The exploitation of mind map topology in qualitative analysis, as a link analysis among nodes, is a novel contribution of this research.

Figure 4.3 The red circles show nodes in the parent level "1," and the green rectangles include nodes in the children level "2," and the blue rectangles show nodes in the grandchildren level "3."

During the conversion of mind maps into text files, a spell-check was carried out, which is important because words are the smallest unit that computer deals with, and it assumes correct and consistent spelling (Stubbs, 1996). Figure 4.4 shows the created text file for the mind map from Figure 4.2.

money financial freedom 1
be in a position for myself and my closest people to live life without worry about its cost 2
being able to achieve one of my goals of giving back to the poor 2
humility 3
gratitude 1
my motivation comes from just being given the chance to live 2
we could have been born a tree lion bug etc but we are a special species that been given the whole world 3
i am just thankful that i even have the opportunity to start a new venture 2
love the challenge 1
always been a competitor 2
facing your fears and doing what makes you uncomfortable will make you grow 2
wisdom 1
the lessons you will learn along the journey to success 2

Figure 4.4 A text file created for the mind map in figure 4.2 shows how nodes are written in sequential order, for example, children nodes will always be under their parents, so a text file can be converted back to a mind map. The numbers at the end of each line represent the level of nodes.

Text file's name composed of three parts. These names are unique for each text file because when the data is read into software, the node in each mind map becomes a separate document with its file's name as an identification number (ID). Thus, nodes belong to the same mind map have the same ID.

The first part is the culture or semester abbreviation letters: F for French, U for U.S., S17 for Spring 2017, and F17 for Fall 2017 and suchlike. The second part is the type of question, i.e., motivation or deterrence, where the motivation was abbreviated with letter M, and the deterrence was abbreviated with letter D. The third part is the student identification letter, which is unique for each student from one dataset.

R, an open-source statistical computing software, was used in this research. Several packages and libraries in R provide valuable functions for text analysis, such as '*tm*' (Feinerer, 2018), '*RTextTools*' (Collingwood et al., 2013), '*quanteda*' (Benoit et al., 2018), '*Tidytext*' (Silge & Robinson, 2016), and '*svs*' libraries. For classification and topic modeling algorithms, '*caret'* (Kuhn, 2008) and *'stm'* (Roberts, Stewart, & Tingley, 2014) were used. For specific tasks, such as separating the level of nodes from the text column, manual functions were written in R.

The datasets, culture and major, were read into R, and a data frame with three columns: serial, ID, and content, was created for each dataset. Figure 4.5 shows a snippet of the data frame that includes the mind map from Figure 4.3. Further data preprocessing is illustrated in the experiments.

Ser	**ID**	**Content**	**Level**
1	S18Md	money financial freedom	1
2	S18Md	be in a position for myself and my closest people to live life without worry about its cost	2
3	S18Md	being able to achieve one of my goals of giving back to the poor	2
4	S18Md	humanity	3
5	S18Md	gratitude	1
6	S18Md	my motivation comes from just being given the chance to live	2
7	S18Md	we could have been born a tree lion bug etc but we are a special species that been given the whole world	3
8	S18Md	i am just thankful that i even have the opportunity to start a new venture	2
11	S18Md	love the challenge	1
12	S18Md	always been a competitor	2
13	S18Md	facing your fears and doing what makes you uncomfortable will make you grow	2
9	S18Md	wisdom	1
10	S18Md	the lessons you will learn along the journey to success	2

Figure 4.5 The first 13 rows of a data frame. The ID values show that these nodes belong to the same mind map, and the map is collected in Spring 2018 for the motivation theme, and the student unique ID is the letter d.

4.4 Quantitative Content Analysis of Mind Maps

Quantitative content analysis of mind maps can reveal more information than that applied to standard text documents, such as the number of nodes per map and map's depth (Beel & Langer, 2011). The map's depth is computed from the hierarchical level of nodes, which equals the sum of multiplying nodes with their levels and dividing that by the number of nodes.

The quantitative analysis demonstrates how students in the data sample utilize mind maps when used as a qualitative data collection tool. The analysis includes statistics of the number of nodes per map, tokens per map and per node, and maps' depth.

The quantitative analysis was performed on the motivation and deterrence subsets within each dataset. It allows for direct comparison between covariates, e.g., U.S. and France, in the same subset, to test if there is any statistically significant variance of how students utilize mind maps. The propositions in this matter are as follows:

- The number of nodes drawn by U.S. and France students varies.
- The number of tokens written by U.S. and France students varies.
- Depth of mind maps drawn by U.S. and France students varies.
- The number of nodes drawn by BUS and CS students varies.
- The number of tokens written by BUS and CS students varies.
- The depth of mind maps drawn by BUS and CS students varies.

These propositions were employed as a quantitative analysis of estimating whether students from groups of interest drew mind maps in similar ways or not. Findings may not be directly related to the process automation, although it can help in conducting preliminary analysis of mind maps, exploring the use of language, expression, way of thinking, and among others between samples of interest. For example, summary statistics of mind maps have been used in communication and linguistic studies to compare learning methods (Jamieson, 2012).

To test for these propositions, first, the Shapiro test for normality was performed on the variables, e.g., number of nodes per map (Shapiro & Wilk, 1965). A one-way Analysis of Variance (ANOAV) was used to test a proposition for a normally distributed variable, while the Mann-Whitney U test “nonparametric” was used to examine a proposition for a non-normally distributed variable (Scheffe, 1999; Nachar, 2008). Results and findings and other summary statistics are presented in Chapter 5.

4.5 Experiment One - The Automated Deductive Analysis

4.5.1 Aim of the Experiment

The first experiment's main goal is to test the hypothesis of automating the deductive qualitative content analysis of mind maps with supervised machine learning, particularly text classification. The experiment explores if the performance of text classification is similar to that of human analysis and whether extracted features based on mind map topology can improve the performance of such process automation. It also includes an entrepreneurship application that tests the feasibility of the framework in EE research.

The rigorous testing for the available approaches to recognize the best solutions for classification tasks has been recommended by Hartmann et al. (2019). The experiment introduces a variety of techniques and methods to achieve better performance, including distinct feature representations of text, potential extracted features, and different classification algorithms. The performance measures applied in this experiment are commonly used in the literature of machine learning applications; the measures of performance are classification accuracy and F-score (Diebold & Mariano, 2002; Labatut & Cherifi, 2012 and Sokolova & Lapalme, 2009).

The deductive approach to content analysis requires the use of prior formulated knowledge and theoretical aspects to connect with the text (Mayring, 2004). As defined in Section 3.2.1, a reference model is used as guidance for manual coding of mind maps. It is the construct that an analyst follows to interpret and classify texts inside nodes. Text classification models need manually labeled text data to train classifiers based on observed patterns (Dumais & Chen, 2000). In this experiment, the following reference model guided the manual analysis.

4.5.2 The Entrepreneurial Process Model

Shane, Locke and Collins (2003) identify and explain how human motivations influence the entrepreneurial process. They built their view on the ground of considering entrepreneurship a creative process, i.e., dynamic, rather than static. They posit that instead of looking at entrepreneurship as a process occurring over time, much of the prior research has looked at it as a profession that certain people choose. Another study by Carsrud and Brännback (2011) argues that motivational differences influence the entrepreneurial process. This research links the students' entrepreneurial motivation to the entrepreneurial process model introduced by William Bygrave as an extension of Moore's (M-B).

Moore's (1986) work intended to integrate past studies of entrepreneurial behaviors and individual characteristics and provide guidance for future research. Moore claimed that his model is different from earlier models because it combines behavior with a detailed stream; the stages for the entrepreneurial process. Moore's model also splits critical attributes, such as locus of control, risk-taking, personal values, job satisfaction, experience, role models, age, and education, by the impact they have on each stage of the process (Moore, 1986).

William D. Bygrave presented an entrepreneurial process model (M-B) that was built on Moore's model. Figure 4.6 shows the stages of the M-B model and the attributes that accompany them. In this enhanced model, a new stage is included, the trigger events. This model connects eight categories, three personal, three environmental, one sociological, and one organizational with one or two of four stages of entrepreneurship. Each category consists of attributes, including the constructs in the Theory of Planned Behavior (Ajzen, 1991).

The M-B process model translates entrepreneurial motivations into the entrepreneurial process model that includes innovation as "initiation," trigger events as "initiation," implementation as "growth," growth as "exist." The responding terms inside quotation marks are defined by Murnieks, Klotz, and Shepherd (2020) as phases of the business development process. The authors state that even though studies of entrepreneurial motivation tend to focus on the business process, they tend to focus on a single phase. They claim that focusing only on one stage has failed to observe a holistic framework for understanding various motives and their influence on the entrepreneurial process. The stage process models also fit the arrangement of topics in the entrepreneurship curriculum (Moroz & Hindle, 2012; Moore, 1986).

The reasons for selecting this model as a reference model are, first, the M-B model includes entrepreneurial motivations defined in the work of (Shane, Locke, & Collins,

2003) including the need for achievement, locus of control, risk-taking, and independence, and more. The B-M model also provides more than 40 attributes related to the entrepreneurial process, which meets the functionality of a model used as a reference to code qualitative data, especially when this data is an answer to what motivates participants to create a new venture.

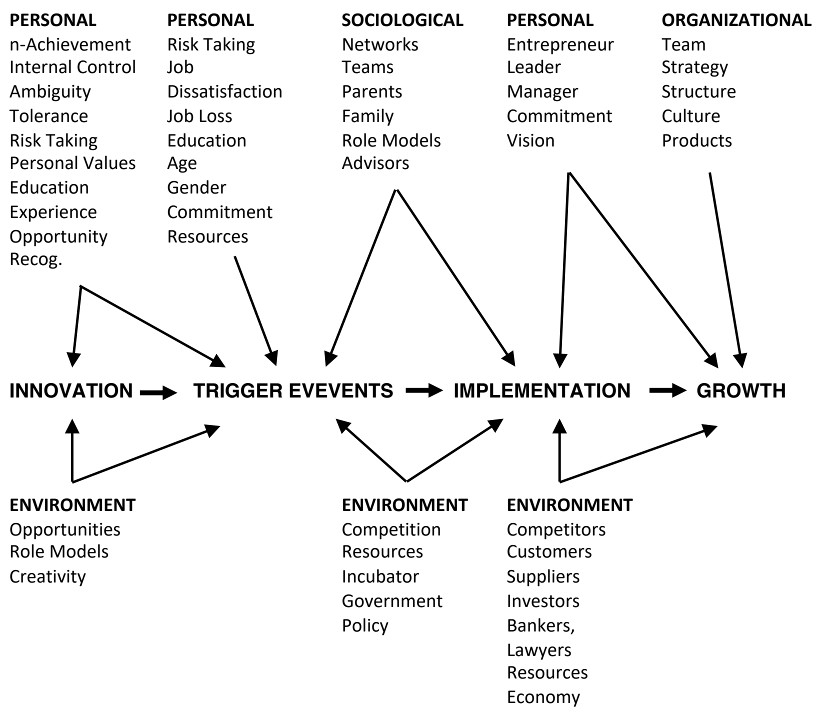

Figure 4.6 The Moore-Bygrave Entrepreneurial Process Model, the M-B model.

4.5.3 Participants

The culture dataset includes a sample of 21 students from School A "U.S." and 26 students from School B "France." Ninety-eight percent of students were senior and junior "3rd or 4th-year" university students, they all lived in the same country where they attended. They all had followed their country's traditional academic program. The students' age was similar, the US mean = 21.0 years and French mean = 20.4 years, and the majority, more than 80%, did not previously work in a start-up company. Government accredited universities allowed all students to take an introductory course to the entrepreneurship course. The two introductory courses, in the U.S. and France schools, were examined and believed to have similar content and learning outcomes, and they were the first undergraduate entrepreneurship courses in the corresponding universities, customarily taken by 3rd-year students. At the time they participated in the study, the students had almost completed the course.

Mind maps were collected from 21 and 26 students from School A, “U.S.” and School B “France,” respectively. To distinguish definite answers from negative ones, students were asked to draw two mind maps, each with a specific central subject that is the problem or question to be solved. The first theme is “Things that motivate me to create a new venture.” The second theme is “Things that deter me from creating a new venture,” focused on positive and negative attitudes, respectively. These themes are referred to as “motivation” and “deterrence” for brevity.

Mind mapping was introduced to all students by their instructors to familiarize themselves with both the software and the process. Students were informed that there were no right or wrong answers. Given that there is a bilingual situation similar to what Lichy and Pon (2015) postulated, and due to the international context of the research, it was essential to conduct the mind mapping exercise in English and French to allow students to respond without limitations of a foreign language. Data collected in French was translated into English. The students in the U.S. used Mindmup.com, an online mind mapping service, while French students used Framindmap.org.

4.5.4 Data Preprocessing for Classification

The initial preprocessing for culture mind maps was introduced in Section 4.3. The dataset has been already uploaded into the software as a data frame. The content column in the data frames, as seen on Figure 4.5, consisting of text was preprocessed by the '*quanteda'* package (Benoit et al., 2018).

The computational preprocessing starts with cleaning texts. Data cleaning is the process of finding incorrect, incomplete, and corrupt inputs of the data and then removing, modifying, and replacing them. In automated text analysis, cleaning handles removing symbols, numbers, punctuations, and stop-words, which have higher frequencies in the English language, e.g., me, we, and are (Manning & Schütze, 1999). Aggarwal and Zhai (2012) state that stop-words are the most frequent words in any investigated corpus, thereby, they can be removed, and the remaining most frequently used words are often the important ones.

4.5.4.1 Feature Representation. The text was quantified as a Document-Term Matrix (DTM) (Bauer et al., 1998). The rows represent documents, and the columns represent terms. Three weighting methods were used, first, TF, as a binary weighting scheme that assigns 1 for occurrence and 0 for no occurrence. Second, TF-IDF, which computes the occurrence of terms in each document and multiplies that with the inverse of occurrence of the same term in all documents. The third method created a document-term matrix but with NPMI weighting scheme. These three main matrices were used to train the classification algorithms.

The Singular Value Decomposition (SVD) was used to lower the dimensionality of the DTM (Golub & Reinsch, 1971). Because a DTM can be a very sparse representation, i.e., it includes many zero inputs, SVD has been a reliable method for extracting the desired low-rank representation from noisy data (Wall, Rechtsteiner, & Rocha, 2003). Extremely too low ranks can also produce noise, and it is crucial to obtain ranks that generate the optimal representation of the main matrix. Three different rank values; 50, 100, and 200 were therefore used to generate optimal representation.

The singular values represent the DTM in terms of a new coordinate system defined by dominant correlations within rows and columns of the matrix (Golub & Reinsch, 1971). Therefore, the singular vectors used for approximation pose a tradeoff between the number of singular values, i.e., ranks, to be computed, and the reconstruction fidelity, i.e., how well the new factorized matrix represents the original one (Erichson et al., 2016).

The feature representations of text generated six different DTMs for each subset: TF, TFIDF, NPMI, SVD-50, SVD-100, and SVD-200.

4.5.4.2 Feature Extraction. The key features extracted from the culture dataset, besides the computation of SVD, were based on mind map topology. The first feature is the level of nodes, which was already inserted into the plain text. From that feature, other features were created. In this study, two methods of extracting features from mind map topology were hypothesized and tested. The first method extracts a feature with an emphasis to the link analysis concept introduced by (Beel & Gipp, 2010). The extracted features were created for each node as follows:

- Feature Branch “L1”: nodes in level 1, i.e., parent nodes, were given unique values, and all nodes under each branch inherit the same value, to distinguish branches inside the same mind map
- Feature L2: nodes in level 2 under one branch root were given unique values, e.g., 1, 2, and 3, to distinguish siblings among children in level 2

- Feature L3: nodes in level 3 under one branch root were given unique values, e.g., 1, 2, and 3, to distinguish siblings among children in level 3
- Feature L4: nodes in level 4 under one branch root were given unique values, e.g., 1, 2, and 3, to distinguish siblings among children in level 4

Table 4.1 shows the extracted features for mind maps in Figure 4.3 following the first method. These features were added to a DTM in a separate classification experiment to measure the effect of mind map topology, under the first method, on the automated analysis performance, which manifests a way of testing hypothesis *H2a.*

Table 4.1 Extracted Mind Map Topology Features Under Method One

Serial	L1	L2	L3	L4
1	a	0	0	0
2	a	1	0	0
3	a	2	0	0
4	a	2	1	0
5	b	0	0	0
6	b	1	0	0
7	b	1	1	0
8	b	2	0	0
9	c	0	0	0
10	c	1	0	0
11	c	2	0	0
12	d	0	0	0
13	d	1	0	0

The second method extracts a feature of mind map topology based on the similar concept presented in Beel and Gipp (2010), but adjusting the link analysis to alternatively includes the following:

- Feature “ID”: nodes inside one map were given unique values, to distinguish one map’s nodes among other nodes in the dataset

- Feature Branch “B”: nodes in level 1 were given unique values, all sub-branch nodes inherit that value, to distinguish branches inside the same mind map

Table 4.2 shows the extracted features for mind maps in Figure 4.3 following the second method. These features were also added to a DTM in another classification experiment to measure the effect of mind map topology, under the second method, on the automated analysis performance, which serves as a way of testing hypothesis *H2b.*

Table 4.2 Extracted Mind Map Topology Features Under Method Two

Serial	ID	B
1	S18Md	a
2	S18Md	a
3	S18Md	a
4	S18Md	a
5	S18Md	b
6	S18Md	b
7	S18Md	b
8	S18Md	b
9	S18Md	c
10	S18Md	c
11	S18Md	c
12	S18Md	d
13	S18Md	d

4.5.5 Training Set

The manual coding for the training set was done first by Bandera et al. (2018b) for 1119 documents, where 582 documents belong to the motivation subset, and 532 documents belong to the deterrence subset. In the primary coding, the documents were manually classified by two professors, reaching a 93% coefficient of reliability, into one of the eight categories of the M-B model shown in Figure 4.6. The second coding done in this research combined the overlapping categories in the M-B model to reduce the target classes into five. Figure 4.7 shows the M-B model's five classes used to classify documents manually.

Figure 4.8 shows the multinomial distribution of the five classes over the motivation subset, and Figure 4.9 shows the multinomial distribution of the five classes over the deterrence subset.

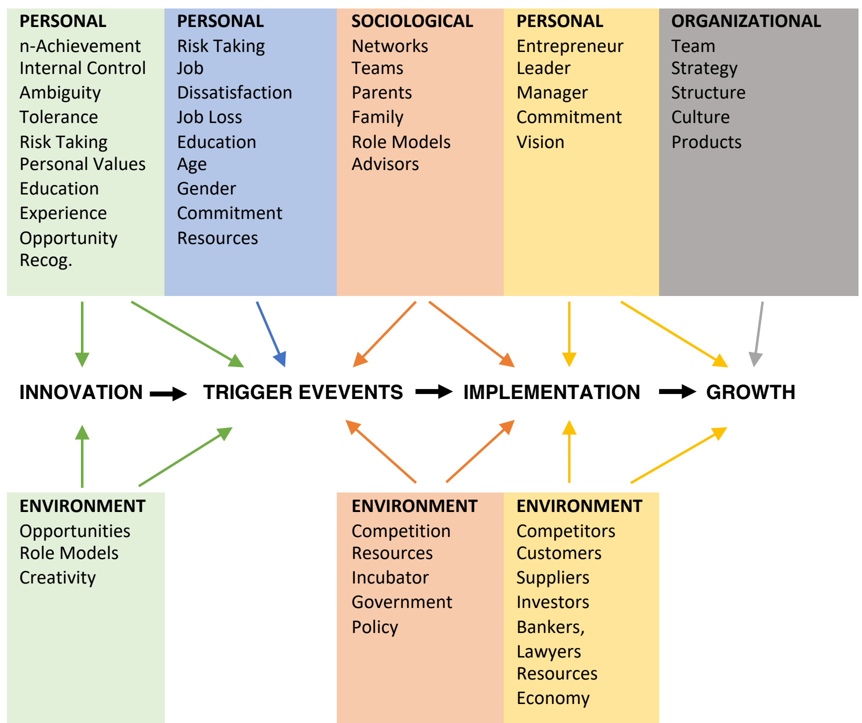

Figure 4.7 The five classes from the M-B Model. The first class, colored by light green, represents personal and environmental attributes that link to Innovation and Trigger Events stages. The second class in light blue stands for personal attributes that link to the Trigger Events stage only. The third class in light orange links attributes of personal and environment to Trigger Events and Implementation stages. The fourth class in golden color represents personal and environmental attributes connected with the Implementation and Growth stages. Finally, the fifth class is in grey and represents organizational attributes that link only to the Growth stage.

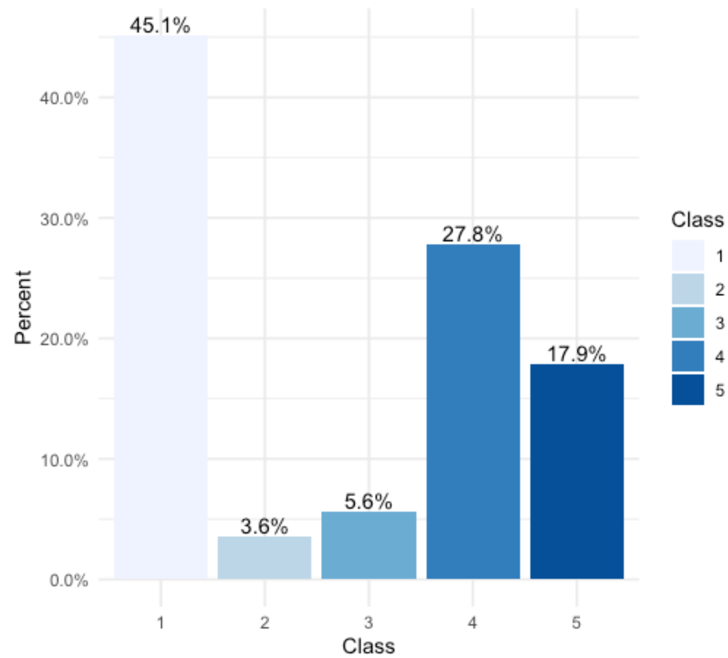

Figure 4.8 Five classes distribution in the motivation set.

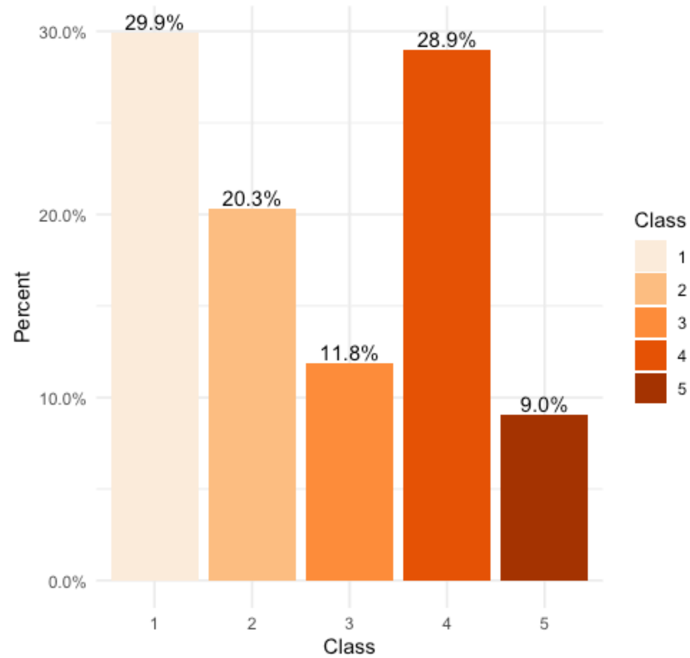

Figure 4.9 Five classes distribution in the deterrence set.

4.5.6 Text Classification (Automated Analysis)

For a text classification task, there are a variety of classification algorithms that can be applied. To seek extensive experimentation, four different classification models, due to their conceptually algorithmic approaches, their proven performance, and their application for text classification, were used (Hartmann et al., 2019). They are Random Forest (RF), Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Logistic Regression (LR).

Random Forest (RF) algorithm has gained much attention in the last years (Hartmann et al., 2019). RF is an ensemble learning method that produces many randomized and uncorrelated decision trees (Breiman, 2001). The decision trees operate to predict a class for one document. The most popular class among all trees becomes the final prediction of the RF classifier. It is a procedure known as bagging (Breiman, 2001). There are many methods to produce randomness for the individual decision trees, for instance, by selecting random features and data subsets (Breiman, 2001). RF overcomes overfitting problems by combining many decision trees based on a heterogeneous randomly drawn subset of features (Domingos, 2012; Sebastiani, 2002).

Random Forest is a robust model against outliers and noise (Breiman, 2001), therefore, it is anticipated to perform highly consistently across datasets. RF can operate automatic feature selection, learn complicated interactions among features, and fit non-linear data (Breiman, 2001). The automatic feature selection of RF has been discussed in the machine learning literature to measure importance of features (Calle & Urrea, 2011). The Mean Decrease Gini (MDG) measure produced by the RF model serves to discover the importance of mind maps topology features when added to the model. According to (Calle & Urrea, 2011), the MDG provides more robust results than the Mean Decrease Accuracy (MDA) measure produced also by the RF model.

When connections are slightly embedded in the text and extend across features, it is believed that RF can handle both content and more complex text classification, where contextual implications need to be understood (Domingos, 2012; Sebastiani, 2002). However, the only downturn of applying RF is that the model running time increases based on the number of decision trees in the ensemble (Breiman, 2001). The solution to that can be a parallelized processing since each tree is operating individually, which allows RF to be computationally efficient. In this experiment, 500 decision trees were used in each run of the model, and a parallelized processing, was implemented.

The second model applied is SVM. It is a discriminative classifier that fits a hyperplane between classes (Scholkopf & Smola, 2001). It was initially built as a binary linear classifier (Cortes & Vapnik, 1995), but through using kernels function, SVM was extended to deal with non-linear problems of higher dimensionality (Scholkopf & Smola, 2001). The Radial Basis Function (RBF), as a kernel function, was accordingly used in this experiment (Schölkopf et al., 1998), see Eq. 4.1. One of the main advantages of SVM is the unlikeliness to overfit, which leads to better generalization (Bennett & Campbell, 2000). The support vectors determine the position of a margin-maximizing hyperplane that separates classes, and a convex optimization problem created by the computation of the margin-maximizing hyperplane parameters can be computationally costly when the sample size and the number of features are so many (Sebastiani, 2002; Moraes, Valiati, & Neto, 2013).

For the multidimensional pair of inputs (x_i, x_j)

$$K(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel)^2 \qquad (4.1)$$

where γ is a parameter for the kernel function.

Support Vector Machine has been proven to provide a useful model for specific text problems such as sentiment and news article classification, and second to its proficiency when dealing with high dimensionality (Bermingham & Smeaton, 2010; Wu et al., 2008). Nevertheless, SVM can be less prone to overfitting, as Joachims (1998) and Pang (2008) postulate. It is expected that SVM could produce comparable results to a simple method like KNN and exceeded by more sophisticated methods like RF.

K-Nearest Neighbor is a lazy learning model or a nonparametric model because it requires no tuning in the training step, and it can be used as a baseline classifier (Yang, 1999; Hu et al., 2016). KNN finds the nearest neighbors of the testing target to assign a class based on most neighbors' classes. The efficiency in that prediction increases as the closeness to neighbors in the same class increases (Yang & Liu, 1999). KNN can compute distances among observations using different techniques; in this experiment, the Euclidean distance function was used because it is the most widely used function in KNN (Hu et al., 2016). KNN, conversely, can be applied to a high-dimensional matrix, the computation of the distances among all training documents becomes costly, especially if the number of documents is large (Sebastiani, 2002). The reason for including KNN in this experiment, besides being a nonparametric method, is the tendency to perform well compared to other models for shorter texts (Aggarwal & Zhai, 2012; Hartmann et al., 2019).

The fourth model is Logistic Regression (LR), and it is frequently applied for classification problems (Kleinbaum et al., 2002). Although the model is primarily used with binary classes, LR can be extended to operate with three or more classes, i.e., multinomial classes (Wright, 1995). The focal mathematical function that underlies LR is the logit function, see Eq. 4.2, which is the natural logarithm of an odds ratio (Peng, Lee, & Ingersoll, 2002). This function is required to transform a continuous number to a probability of success, which then allows the regular linear regression model, see Eq. 4.3, to deal with certain classes (Peng, Lee, & Ingersoll, 2002). LR model is capable of producing high prediction accuracy for complex applications such as text classification (Genkin, Lewis, & Madigan, 2007).

$$P(x) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} \tag{4.2}$$

$P(x)$ is used instead of $E(Y|x)$ to simplify notation, the parameters β_0 and β_1 denotes the slope of a linear function and the coefficients of x (Hosmer Jr, Lemeshow, & Sturdivant, 2013)

$$E(Y|x) = \beta_0 + \beta_1 x \tag{4.3}$$

a linear equation in regression, it is read as the expected value of *Y,* given the value of *x.*

Logistic Regression suffers from computational efficiency while fitting the model, also, while using the fitted model on an unseen set, there can be a computational problem. The LR model is also prone to the overfitting problem (Genkin, Lewis, & Madigan, 2007).

Regularization is a method used in machine learning to avoid the overfitting problem, especially when there is only a small number of training examples (S.-I. Lee et al., 2006). In this experiment, the logistic regression model has been tuned to find the best regularization method. The cost constraints violation, which is the tradeoff between regularization and correct classification on the training set, was tuned as a hyperparameter for the model (Fan et al., 2008). A tolerance of termination criterion for optimization was also tuned for the model (Fan et al., 2008). The best combination of these hyperparameters was used in the final model.

The four classification models have been applied to the six different matrices for each group of the datasets. Each text classification model's goal was to develop a similar interpretation of the text as a human coder. Each dataset was divided into a training set, 90% of the data, and a testing set, 10% of the data to obtain an unbiased estimate of out-of-sample accuracy (Hartmann et al., 2019). Five-fold Cross-Validation (cv) was used to tune the most efficient hyperparameters for RF, SVM, and LR models on the training set (Arlot & Celisse, 2010), while the KNN model did not use this. Each training set was partitioned into five equal training and validation subsets, k = 5, and the process repeated three times $j = 3$.

The grid search procedure was essential to test a variety of hyperparameter weights to identify the best of them based on validation accuracy, which is computationally complex. Hyperparameter tuning is essential, however, since default weights might not be efficient enough for selected models. The hyperparameter weights that achieved the best accuracy across all the 3x5 cv were used to fit the final models. The 3x5 cv also gives an estimate of the variance of the algorithms (Salzberg, 1997).

The four classification algorithms used in this experiment have been selected to cover different approaches to classification. LR introduces the linear approach with tuning for best regularization (Freedman, 2009). The non-linear approach is performed by SVM with an RBF kernel (Schölkopf et al., 1998). RF represents the ensemble approach, the bagging method, that includes a multitude of decision trees (Breiman, 2001). KNN represents the nonparametric approach (Yang, 1999). Many R packages, including "*caret*" and "*RandomForest,*" have been used for running the models (Kuhn, 2008).

4.5.7 Measures of Experiment One

Testing the hypothesis of automating deductive analysis for mind maps with classification algorithms was performed through measuring if the accuracy of a given model is better than the no-information rate with a 95 percent confidence interval (Read & Cressie, 2012). A multinomial test with one-side, because it is only concerned about being better than chance, was used (Read & Cressie, 2012). This procedure of testing the significance of a classification task followed the work of (Salzberg, 1997). Because the baseline for the test was the human analysis, i.e., naïve classifier based on the multinomial distribution of manual classification, the significance of a new classifier can be compared to the performance of human analysis (Read & Cressie, 2012).

The classification performance measures include prediction accuracy when a class distribution was balanced and F-score when classes were unbalanced to eliminate a possible bias towards dominant classes (Sokolova & Lapalme, 2009). The accuracy of a model is the sum of all accurate predictions on the hold-out test set divided by the sum of all predictions, explained in Section 3.5 (Diebold & Mariano, 2002; Labatut & Cherifi, 2012). F-score is the average of multiplying recall and precision for all classes (Sokolova & Lapalme, 2009). All accuracies and F-score reported in this research were based on the models' prediction on the testing set, see (Preoţiuc-Pietro, Lampos, & Aletras, 2015) for a similar approach. The reported p-value of the multinomial test was also based on the models' prediction on the testing set. To avoid redundancy, only the p-value for the best model across the same dataset is reported.

For testing the hypothesis of mind map topology effect on the improvement of a classification task, the Wilcoxon signed-rank test was used (Wilcoxon, 1992), as an alternative nonparametric test to the paired t-test. The paired t-test for comparing two classifiers has been widely used (Yu et al., 2014). The assumptions of the paired t-test might be violated because the sample of observations were less than 30, and so normality test may not be accurately reported (Demšar, 2006; Salzberg, 1997).

The Wilcoxon signed-rank test is a recommended alternative, which can be even more potent than the t-test (Demšar, 2006). The Wilcoxon signed-rank test ranks the differences in two classifiers' performances for each dataset and compares the ranks for the positive and the negative differences (Wilcoxon, 1992). Besides, the classifiers' performance was measured using F-score over multiple datasets. The procedure followed the guidelines presented by Demšar (2006) for comparing two classifiers over multiple datasets, where each dataset represents a sample item.

Classifier A was applied to a DTM, and classifier B was applied to the DTM that also included the features of mind map topology. The two classifiers were trained with 2x10 cv, i.e., ten folds repeated twice, and tested on an unseen dataset. The algorithm used for both classifiers was the regularized logistic regression "LR" (S.-I. Lee et al., 2006) because it is one of the standard algorithms used for a classification task (Kleinbaum et al., 2002). The cross-validation tuned the classifier's hyperparameters before seeing the test set for both classifiers, as recommended by Salzberg (1997). When the hyperparameters appeared to be optimal, F-score can finally be measured on the test data. The two classifiers were applied on the motivation and deterrence subsets of the culture dataset with different feature representations, i.e., TF, TF-IDF, NPMI, and SVD with different ranks. The same testing procedure was performed twice, one for each method of feature extraction of mind maps features.

The feature selection of RF with the Mean Decrease Gini (MDG) score is reported as an additional measure of examining the importance of mind map topology features among all other features (Calle & Urrea, 2011). The measure is reported for the two methods of extracting mind map topology features. This measure was not intended to test the hypotheses, but to serve as further analysis of the importance of mind map topology. It was performed separately using RF as a classifier and it was only applied to a dataset that included mind map topology features.

4.5.8 Entrepreneurship Education Application

This EE application is intended to demonstrate the feasibility and applicability of the automation framework into the EE research, not only from the perspective of the process automation but also from the perspective of utilizing such a mechanism to measure students' entrepreneurial motivation and making valid statistical inference.

The application attempts to identify how cultural norms can affect entrepreneurial motivations of students enrolling in a similar curriculum of entrepreneurship but belonging to two different cultures, the U.S., and France.

Research into the effect of culture on entrepreneurship outcomes is essential for many reasons. For instance, Bandera et al. (2018b) state that because of the globalization of entrepreneurship education, cultural disparities should be understood by instructors when creating and modifying the curriculum.

Cultural theories have been used in EE research to compare students of diverse cultures (R. S. Shinnar, Giacomin, & Janssen, 2012). One of the theories used in similar studies is Hofstede's theory of National Culture (Hayton, George, & Zahra, 2002; Hofstede et al., 2004). The reasons for selecting Hofstede's theory over others include the broad applications of Hofstede's dimensions within many business disciplines, such as marketing, international business, and project management (Soares, Farhangmehr, & Shoham, 2007; Hofstede, 1983; Hofstede, 1994). Hofstede's theory also provides direct measures that can be used to make a comparison between students from different cultures (Mueller & Thomas, 2001).

The dimensions of Hofstede's theory are Power Distance Index (PDI), Individualism (IDV), Masculinity (MAS), Uncertainty Avoidance Index (UAI), Long-Term Orientation (LTO), and Indulgence (IVR) (Hofstede, 2003). Hofstede’s work has identified 76 countries in a score on a scale that runs from 0 to 100 for every dimension (Hoftede, Hofstede & Minkov, 2010). Table 4.3 demonstrates how the six dimensions define attributes based on low and high scores (Hofstede, 2003). These scores have helped to study the dissimilarities and likeness between students and individuals’ entrepreneurial outcomes among cultures, as described by Bandera et al. (2018b), Hofstede et al. (2004), Shinnar, Giacomin and Janssen (2012).

Table 4.3 Attributes linked with Hofstede's Culture Dimensions

	Low Scores Attributes	**High Scores Attributes**
Power Distance	Power is expected to be distributed equally. Education is student-centered.	Power is expected to be distributed unequally. Education is teacher-centred.
Individualism	We ‘mind’. Ties between individuals are substantial.	I ‘mind.’ Ties between individuals are loose.
Masculinity	Quality of life was more important than quantity. Family is more important than work.	Quantity of life was more important than quality. Work is more important than family.
Uncertainty Avoidance	More tolerant toward ambiguity and uncertainty	Stricter toward ambiguity and uncertainty
Long-Term Orientation	Efforts should bring quick results. Success and failure are luck	Perseverance is what matters. Success and failure are efforts
Indulgence	Strict social norms suppress gratifications	Enjoying life and having fun is freely allowed

Hofstede, G. (2003). Cultural dimensions. *www. geert-hofstede. com*. “*accessed on December, 2nd 2019*”

For example, Mueller and Thomas (2001) argue that an inclination to innovation is more prevalent in low uncertainty avoidance cultures, U.S. as an example, than in high uncertainty avoidance cultures, France as an example. Similarly, they posit that in cultures with high individualism score, an internal locus of control orientation is more prevalent than in cultures with a low score, collectivistic cultures. Mueller and Thomas (2001) also attach innovativeness combined with an internal locus of control to low uncertainty avoidance and individualistic cultures than in high uncertainty avoidance and collectivistic cultures. Their findings conclude that, for entrepreneurial thinking, culture may generate distinctions across regional boundaries and national. EE should consider this relationship between culture and entrepreneurial outcomes by helping students change their focus toward creativity, self-reliance, independent action, and flexible thinking (Mueller & Thomas, 2001).

S. M. Lee and Peterson (2000) also show that low scores for the dimensions of uncertainty avoidance, power distance, and a long-term orientation, along with high scores for the individualism dimension, make a more robust entrepreneurial outcome.

Table 4.4 Scores for the US and France on Hofstede's Dimensions

	PDI	**IDV**	**MAS**	**UAI**	**LTO**	**IVR**
The U.S.	40	91	62	46	29	68
France	68	71	43	86	63	48

Hofstede, G. (2003). Cultural dimensions. *www. geert-hofstede. com*. "*accessed on December, 2nd 2019*"

Table 4.4 shows the Hofstede's dimensions scores for the U.S. and France. The dissimilarities between the two cultures favor the U.S. in terms of entrepreneurial motivations (Hofstede, 2003). The Global Entrepreneurship Monitor GEM, which is one of the most extensive studies of entrepreneurship in the world, supports this finding by reporting that U.S. holds a lower rate of fear of failure and a higher rate of early-stage entrepreneurial activity (Singer, Amorós, & Moska, 2015). Boissin et al. (2009), on the contrary, indicate that in France, an entrepreneur is seen as a boss of small to the medium enterprise rather than a dynamic individual who should consistently innovate, while in U.S., an entrepreneur is related more to risk-taking, innovation, and dynamism.

Some studies conclude that the association between entrepreneurial motivations and culture might be contradictory (Verheul et al., 2002). For instance, Baum et al. (1993) compare entrepreneurs and non-entrepreneurs and find that high uncertainty avoidance, high power distance, and low individualism surprisingly associated with entrepreneurial intentions. They presume that dominant cultural values may cause dissatisfaction among people that consequently make them entrepreneurs.

Hofstede's national culture dimensions and GEM analytics show that France has a higher score of uncertainty avoidance and long-term orientation in which the M-B model links it with the first stage of entrepreneurship, Innovation. Thus, this research includes the following case problems:

- U.S. students are more inclined to the Innovation Stage of the M-B model than French students do
- French students show more apprehension to the Innovation Stage of M-B model than U.S. Students do

On the other hand, Hofstede's model presents that France cultural norms is more tolerant than U.S. norms in terms of power distance, which entails more accepting of an organizational hierarchy (Hofstede et al., 2004). The M-B model links this to the last stage of the entrepreneurship process, Growth. The research comprises the following case problems:

- French students are more inclined to the Growth Stage of M-B model than U.S. students do
- U.S. students have more apprehension towards the Growth Stage of M-B model than French Students do

The case problems were tested after the conversion of the five classes, used to classify nodes, to a four-component vector for each node w_d, where each component is the node's weight on the s^{th} stage of the M-B model, i.e., Innovation, Trigger Event, Implementation, and Growth. The conversion is necessary because, first, case problems have used M-B's four stages as measurement variables. Second, the five classes are shared between the four stages, as seen in Figure 4.7. The conversion allows for direct measure of the four stages. Generating the four-component vector for each node was carried out as follows:

- A four-component vector for a node is w_d = [s_1, s_2, s_3, s_4]
- Class 1: gives .5 to both s_1, s_2 "Innovation and Trigger Event's weights," e.g., w_d = [.5, .5, 0, 0], because attributes of class 1 are shared between the two stages
- Class 2: gives 1 only to s_2 "Trigger Event's weight," e.g., w_d = [0, 1, 0, 0], because attributes of class 2 are only connected to stage 2, Trigger Event
- Class 3: gives .5 to both s_2, s_3 "Trigger Event and Implementation's weights," e.g., w_d = [0, .5, .5, 0], because attributes of class 3 are shared between the two stages
- Class 4: gives .5 to both s_3, s_4 "Implementation and Growth's weights," e.g., w_d = [0, 0, .5, .5], because attributes of class 4 are shared between the two stages
- Class 5: gives 1 only to s_4 "Growth's weight," e.g., w_d = [0, 0, 0, 1], because attributes of class 5 are only connected to stage 4, Growth

The mean of a mind map was computed as follows:

- Mind map mean:

$$w_m = [\sum_{d=1}^{i} \frac{s_{i,1}}{s_{i,1:4}}, \sum_{d=1}^{i} \frac{s_{i,2}}{s_{i,1:4}}, \sum_{d=1}^{i} \frac{s_{i,3}}{s_{i,1:4}}, \sum_{d=1}^{i} \frac{s_{i,4}}{s_{i,1:4}}] \qquad (4.4)$$

where $\sum_{d=1}^{i} s_{i,1}$ is the sum of stage one weights in all nodes inside mind map m, $\sum_{d=1}^{i} s_{i,2}$ is the sum of stage two weights, and the same for $\sum_{d=1}^{i} s_{i,3}$ and $\sum_{d=1}^{i} s_{i,4}$. In the denominator, $\sum_{d=1}^{i} s_{i,1:4}$ is the sum of the four stages' weights in all nodes inside mind map m. Hence, w_m represents the average weights of the M-B stages inside m.

This method of computing mind maps' weight:

- Gives nodes inside maps an equal chance to contribute to the total weight of a mind map
- Prevents bias that might be caused by the number of nodes for each student. For example, it is found that the US students drew more nodes than French students
- Allows attributes of the M-B model to be shared by their stages

The case problems can be mathematically expressed as follows:

- *Case Problem* $_{1a}$: $s_1|_{\text{US, motivation}} > s_1|_{\text{France, motivation}}$
- *Case Problem* $_{1b}$: $s_1|_{\text{France, deterrence}} > s_1|_{\text{US, deterrence}}$
- *Case Problem* $_{1c}$: $s_4|_{\text{France, motivation}} > s_4|_{\text{US, motivation}}$
- *Case Problem* $_{1d}$: $s_4|_{\text{US, deterrence}} > s_4|_{\text{France, deterrence}}$

The probability distribution of s_1 and s_4 of the motivation and deterrence mind maps were tested for normality by the Shapiro test, which is a general normality test (Shapiro & Wilk, 1965). The test rejects the hypothesis of normality when the p-value is less than the alpha level of 0.05. For a normally distributed weight, a linear regression analysis (Kutner et al., 2005) was used as a statistical test to measure the variance between the two cultures. Weights were treated as the outcome variable and culture as the independent variable. For a non-normally distributed weight, the Mann-Whitney U test (Nachar, 2008), a nonparametric test, was used as a statistical test to compare the weights of U.S. and France.

4.6 Experiment Two – The Automated Inductive Analysis

4.6.1 Aim of The Experiment

This experiment aims at testing the hypothesis of automating the inductive qualitative content analysis of mind maps with the Structural Topic Model (STM) (Roberts et al., 2013). As the comprehensive testing for the available methods to pick out the best solution for the automation task is recommended, the experiment also compares the STM's performance to the standard LDA's performance for automating the inductive analysis of mind maps (Hartmann et al., 2019; Blei, Ng, & Jordan, 2003). The comparison with LDA followed a similar approach to that presented by Roberts et al. (2014).

As in experiment one, an EE application is presented as a way of testing the capability of using STM as a topic modeling algorithm in EE studies. The EE application is operated on the second dataset "major" to deduce the effect of a student's college on her entrepreneurial motivation.

4.6.2 Participants

In testing the process automation of STM to automate inductive analysis, the culture dataset was used, in which participants were already described in Section 4.5.3. On the other hand, for the STM application into EE problems, a new dataset was used, the "major" dataset.

Surveys were administered to all students in five sections of an introductory entrepreneurship course, with a similar curriculum and have been taught by the same instructor. The courses were taught between the Spring 2017 and Spring 2019 semesters. 226 complete mind maps were collected from 113 students. Most of the students were seniors and junior students, i.e., third or fourth-year university students. The students belonged to different academic colleges; out of the 113 students, there were 38 students from the School of Management “BUS” and 42 students from the School of Computing “CS.”

The mind mapping concept was introduced to students, and they were required to familiarize themselves with both the software and the process. Students were informed that there were no right or wrong answers. Students were asked to draw two mind maps, one intended to reflect positive opinions, and the other asked about negative ones. The first mind map contains a central subject of “Things that motivate me to create a new venture,” The second mind map includes a central subject of “Things that deter me from creating a new venture.” The students used Mindmup.com to draw mind maps. Then, they used a variety of formats to deliver their mind maps. For example, they used PDF, screenshots, and Microsoft Word.

4.6.3 Data Preprocessing for STM

The ‘*stm’* package in R offers a particular function, *textProcessor*, that reads the text directly from the data frame and apply all preprocessing functions at once, including converting texts to lower case, removing punctuation, removing stop words, removing numbers, and then creating an output (Roberts, Stewart, & Tingley, 2014). The function *prepDocuments* was then used to build three components for the model: documents, vocab, and metadata. Documents refer to a list of word indices and their counts, while vocab is a numerical vector representing the words in all documents. A metadata matrix differentiates STM from other topic modeling algorithms (Roberts, Stewart, & Tingley, 2014). Metadata is the document's covariate, where culture, U.S. and France, was the documents' covariate in the process automation experiment, and major, BUS and CS, were the documents' covariate for the EE application.

For illustration purposes, Figure 4.10 shows how two documents appear after preparing them with *prepDocuments* function with R nomenclature to delineate documents, rows, and columns. The first document consists of five words, in the column, and the rows represent the words' indices and frequencies. The second document includes three words.

```
[[1]]
      [,1] [,2] [,3] [,4] [,5]
[1,]    21   23   87   98  112
[2,]     2    1    1    1    1

[[2]]
      [,1] [,2] [,3]
[1,]    16   61   90
[2,]     1    1    1
```

Figure 4.10 Documents representation in STM.

4.6.4 STM (Automated Analysis)

The intractability underlying the computation of topic models requires external analysis to understand any distinct tradeoffs between competing parameters (Roberts, Stewart, & Tingley, 2014). Therefore, the STM package in R includes a variety of measures that can be used to assess the quality of the model, such as the analyst's choice of the number of topics.

Structural Topic Model is an unsupervised method and the number of topics needs to be determined before running the model. This experiment applied two methods for selecting the best value of *k*, i.e., the number of latent topics to be discovered (Roberts, Stewart, & Tingley, 2014). The first method was introduced by Mimno et al. (2011); the method can be implemented by setting the initialization type to "Spectral" and K to 0 when running the STM function in R. The main concept of the spectral initialization is to find the vertices of the convex hull of the word co-occurrences (Roberts, Stewart, & Tingley, 2014). In solving the convex hull, the matrix was decomposed by the algorithm into a low dimensional space using t-distributed stochastic neighbor embedding (Van Der Maaten, 2014). The decomposition of the matrix presents the advantage of automatically selecting the number of topics. However, this process does not estimate the actual number of topics. Instead, it can be helpful to approximate it. It holds a computational advantage because it only needs to be run once (Roberts, Stewart, & Tingley, 2014).

The second method evaluates a selected value of *k* by using the function *searchK* from the "*stm"* package in R, this represents a data-driven approach (Roberts, Stewart, & Tingley, 2014). The default initialization is the spectral initialization because of its stability. This function runs several automated tests to assess the number of topics by computing the held-out likelihood and residual analysis, aside from calculating the average exclusivity and semantic coherence for all topics (Taddy, 2012; Wallach et al., 2009).

The second method allows for direct comparison among a range of *k* values, where the metrics of semantic coherence and exclusivity were considered the most important scores in this experiment. The reasons for giving semantic coherence and exclusivity more value than residual analysis and held out likelihood are, first, found in this study that residual analysis score is getting near zero as *k* value increasing, regardless of any other factors as the semantical meaning of topics. Table 4.5 shows how residual analysis decreases as the value of *k* increases. Second, Chang et al. (2009) conclude that topic models selected based on held-out likelihood may infer less semantically meaningful topics.

Semantic coherence is presented by Mimno et al. (2011) and it is related to pointwise mutual information (Newman et al., 2010). The metric is maximized when the highest probability words in a topic frequently co-occur together. Mimno et al. (2011) demonstrate that human judgment of topic quality agrees well with this metric. A study by Roberts et al. (2014), conversely, noticed that getting high semantic coherence can be achieved if quite familiar words dominate a few topics. Thus, they prefer to include the exclusivity measure to mitigate semantic coherence of high probability words inside topics.

This experiment used the co-occurrence between words, i.e., semantic coherence, and exclusivity of words to a given topic to measure the quality of topics.

4.6.4.1 Evaluating *k*. Table 4.5 demonstrates a snippet of the second method results for the motivation subset, after testing *k* values from 3 to 60. From the results, the highest combination of semantic coherence and exclusivity was found in *k* = 36, 37, 38. Additional analysis for these values, which serves as a diagnostic test for the model, was performed as recommended by Chang et al., 2009 and Roberts, Stewart and Tingley (2014), including plotting of the models' topic correlation and most FREX words (Figure 4.11). Table 4.6

exhibits the results and the value of *k* was set to be 36 for STM applied to the motivation subset.

Table 4.5 The First Searching for *k* - Motivation

K	Exclusivity	Sem-Coh	Held-out	Residual
3	8.080	-266.246	-5.783	6.757
5	8.874	-259.628	-6.072	4.415
10	9.416	-244.655	-5.869	5.911
15	9.552	-254.297	-5.838	3.219
20	9.716	-241.118	-5.767	9.892
25	9.783	-237.520	-5.735	-7.673
30	9.751	-237.992	-6.046	-2.939
35	9.756	-236.380	-6.288	-1.476
36	**9.843**	**-225.810**	-5.359	-1.367
37	**9.851**	**-222.592**	-5.680	-1.010
38	**9.860**	**-224.334**	-5.573	-0.854
39	9.786	-241.057	-6.271	-0.903
40	9.787	-243.447	-5.796	-0.942
45	9.779	-239.689	-5.921	-0.716
50	9.781	-247.118	-6.529	-0.539
55	9.762	-234.459	-6.465	-0.357
60	9.809	-23.554	-6.461	-0.348

Table 4.6 The Second Searching for k Results

K	Exclusivity	Sem-Coh
35	9.765	-243.265
36	**9.764**	**-234.855**
37	9.748	-237.841
38	9.778	-244.619
40	9.754	-241.731
45	9.787	-241.048

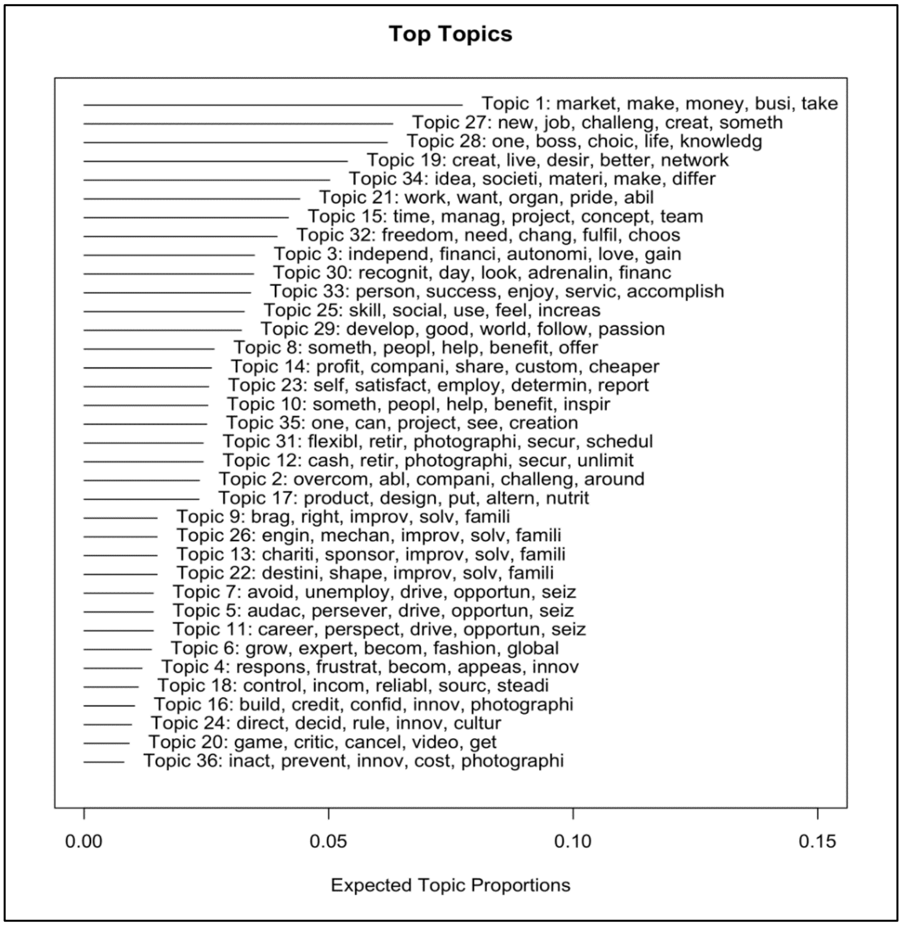

Figure 4.11 STM topic probability in the motivation subset with the most five execlusive words in each topic.

For the deterrence subset, Table 4.7 shows the results of testing different values of *k*. As seen on the table, the values from 30 to 36 hold the best semantic coherence and exclusivity. Additional analysis that includes plotting of the models' topic correlation and most FREX words, as shown on Figure 4.12, was performed on values between 30 to 37. Table 4.8 shows the results and the best value of *k* was found to be 34.

Table 4.7 The First Searching for *k* - Deterrence

K	Exclusivity	Sem-Coh	Held-out	Residual
3	8.105	-247.129	-5.453	6.508
5	8.792	-259.583	-5.866	4.181
10	9.328	-256.539	-5.501	3.230
15	9.648	-231.398	-5.026	6.633
20	9.720	-235.856	-5.473	13.689
25	9.769	-239.339	-5.190	-6.530
30	**9.811**	**-218.423**	-5.545	-2.875
31	**9.815**	**-216.062**	-5.527	-2.416
32	**9.820**	**-215.751**	-5.515	-1.521
33	**9.827**	**-222.200**	-5.544	-1.801
34	**9.832**	**-215.420**	-5.710	-1.224
35	**9.835**	**-221.195**	-5.515	-1.194
36	**9.843**	**-215.268**	-5.581	-1.142
37	9.691	-241.660	-6.066	-1.192
38	9.841	-224.211	-5.286	-0.823
39	9.845	-221.866	-5.350	-0.987
40	9.805	-228.990	-6.242	-0.640
45	9.754	-236.228	-6.837	-0.500
50	9.754	-235.743	-6.499	-0.382
55	9.780	-232.139	-5.715	-0.376
60	9.787	-226.578	-6.145	-0.266

Table 4.8 The Second Searching for k Results - Deterrence

K	Exclusivity	Sem-Coh
30	9.812	-218.067
31	9.820	-220.885
32	9.820	-220.667
33	9.819	-223.086
34	**9.846**	**-214.855**
35	9.840	-215.026
36	9.838	-211.675
37	9.838	-222.839

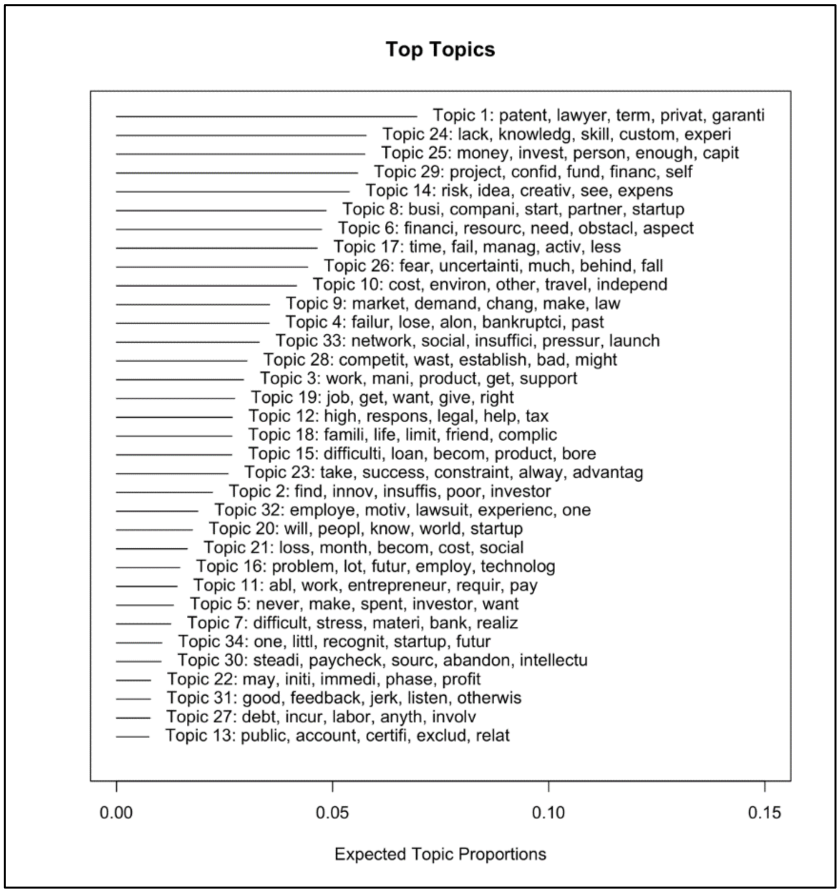

Figure 4.12 STM topic probability in the deterrence subset with the most five execlusive words in each topic.

Finally, STM for the motivation subset includes the covariate of culture, U.S. and France, to estimate topic prevalence with 36 topics, 587 documents, and a 534-word dictionary. STM for the deterrence subset includes the same covariate of culture to estimate topic prevalence with 34 topics, 532 documents, and a 509-word dictionary.

STM generates θ_d and β_k, where the former returns the probability of topics over documents, and the latter shows the probability of words over topics. θ_d is a primary finding obtained from STM, it exposes the probability of k latent topics discussed in each document.

4.6.5 Measures of Experiment Two

The assessment of the STM topics in this experiment followed a similar approach to that presented by Gadarian and Albertson (2014), Roberts et al. (2014), and Baumer et al. (2017). The authors recommend comparing STM topics against a more robust analysis of the same documents.

The manual analysis for this comparison was taken from the work of (Bandera et al., 2018b), where motivation nodes were manually assigned into 31 attributes, and deterrence nodes were assigned into 34 attributes from the M-B model (Figure 4.6), it is referred to them as “categories.” The coefficient of reliability of the manual assignment of categories reached 93%, representing a robust and reliable analysis that can be used for such comparison (Landis, Koch, 1977). Using the manual analysis from Bandera et al. (2018b) as a baseline for the comparison with STM ensures the experiment's internal validity (Burla et al., 2008). Table 4.9 shows the frequency of categories in the motivation and deterrence subsets, they refer to attributes from the M-B model, see Figure 4.6.

The procedure of statistically comparing STM latent topics and a robust analysis of the same documents has not been clearly defined (Baumer et al., 2017). Most of the comparisons between manual analysis and topic modeling in the comparative studies are topic-specific and not comprehensive, and no statistical testing for such comparison has been used (Hagen, 2018). For instance, Roberts et al. (2014) compared only specific topics to the manual analysis; the authors report that documents with high proportions of topic 1 are more likely to be categorized as fear and anger and are rarely categorized as enthusiasm or not categorized.

Table 4.9 Frequency of Manual Categories: Motivation - Deterrence

Category - M	**Freq**		**Category - D**	**Freq**
Personal Values	81		Resources	91
n-Achievement	65		Risk Taking	57
Leader	64		Competitors	32
Op. Recog.	61		Networks	30
Vision	38		Investors	29
Strategy	36		Op. Recog.	29
Products	28		Experience	20
Experience	23		Family	20
Structure	23		Internal Control	19
Networks	18		Customers	18
Customers	17		Manager	18
Manager	16		Strategy	14
Opportunities	16		Commitment	13
Commitment	15		Gov. Policy	13
Team	12		Products	13
Entrepreneur	11		Team	13
Internal Control	10		Ambiguity T.	11

In this experiment, a two-step measure was carried out to compare between STM and the manual analysis. The first step was concerned with measuring the association between an STM (θ_d) and a manual category (C_d), both assigned to the same document *i*. This explains whether the STM assignment for latent topics persists with the manual assignment of category across documents. The second step was performed to measure the similarity between STM (β_k), i.e., words probability over topics, with words probability over manual categories.

For the first step, Pearson's Chi-square test of independence was applied (Greenwood & Nikulin, 1996). To ensure validity of the Chi-square testing of independence, categories or latent topics with a frequency less than 20 were removed

(McHugh, 2013). The assumptions of the Chi-square test of independence between the two categorical variables, i.e., manual categories and latent topics, have been tested (McHugh, 2013). Besides, the p-value for all tests was computed with a Monte Carlo test with 10000 replicates (Hope, 1968).

Table 4.10 demonstrates a sample of documents and their manual categories and dominant topics found by STM.

Table 4.10 Manual Categories and Dominant Latent Topics Assigned to a Sample of Documents

Document ID	**Category**	**Dominant Topics**
FMa	n Achievement	Topic.19
FMa	Manager	Topic.15
FMa	Opportunity Recog	Topic.8
FMb	Vision	Topic.6
FMb	Leader	Topic.1
FMb	Manager	Topic.30
FMb	Leader	Topic.3
FMb	Manager	Topic.21
FMb	Strategy	Topic.21
FMc	Leader	Topic.3
FMc	n Achievement	Topic.19
FMg	Opportunity Recog	Topic.8

The second step measures the similarities between an STM topic and a manual category, both assigned to the same document, in terms of words probability inside them. In STM, β_k represents the probability of words over topics, and a manual β_c was computed comparably to β_k to generate word probability over manual categories. However, the β_k

of STM includes all words in the corpus. On the other hand, β_c includes only certain words for each category, because it is built directly from documents assigned to that category, thus, containing a variety of lengths. Table 4.11 shows the most frequent categories and their words count.

To solve the different numbers of words in β_k and β_c, only the top 88 and 73 words in β_k for the motivation and deterrence subsets were held. The limits of 88 and 73 were chosen to match the maximum number of words inside the motivation and deterrence manual categories, as seen in Table 4.11. By eliminating less important words probability from β_k, the bias in calculating word-level similarity has been decreased.

The computation of similarity was carried out by cosine similarity, as one of the most popular similarity measures applied in text analysis (Huang, 2008). When groups of words, e.g., latent topics and manual categories, are represented as term vectors, e.g., β_k and β_c, the similarity of two groups can be inferred from the correlation between their term vectors (Huang, 2008). An essential property of the cosine similarity is its independence of group length (Huang, 2008).

The similarity between an STM latent topic β_k and a manual category β_c assigned to the same document is:

$$SIM_d\ (\beta_k, \beta_c) = \frac{\beta_k \cdot \beta_c}{|\beta_k| \times |\beta_c|} \tag{4.2}$$

The computation of cosine similarity for the β_c with each β_k of the top three topics assigned by STM was performed separately. The three top topics in each document were assumed to represent most of the STM θ_d in each document.

Table 4.11 Categories with Most Words – Motivation and Deterrence

M-Category	Words Count	D-Category	Words Count
Personal Values	88	Resources	73
Opportunity Recog.	78	Investors	46
n-Achievement	68	Networks	42
Strategy	68	Risk-Taking	41
Vision	56	Commitment	36
Leader	47	Oppo Recog.	36
Structure	35	Competitors	35
Experience	29	Customers	29
Opportunities	27	Family	26
Entrepreneur	23	Manager	26
Customers	21	Strategy	25
Networks	21	Inter Control	23
Manager	20	Team	23
Team	20	Gov Policy	20

For testing the benefit of treating nodes as the unit of analysis when performing STM, hypothesis 2_c, the average of the dominant topics probability given by θ_d across all documents was computed to check if nodes were assigned to as close as one topic. The domination was determined to exist if the average of the dominant topics probability was statistically more significant than the null estimate, $\hat{p} = .5$. It means that a dominant topic holds as more probability per node as the sum of all remaining topics probability.

An exact binomial test was used for testing the domination significance (Agresti & Coull, 1998). The parameters for the test are the probability of success r, which equals:

$$r = \frac{\#\ of\ documents\ exceeding\ the\ null\ estimate\ of\ \hat{p}\ =\ .5}{Total\ number\ of\ documents} \quad (4.5)$$

and the number of observations "n". All the test's assumptions were tested, and the probability of success and failure was dichotomous, "success" for topic probability > .5 and "failure" otherwise.

4.6.5.1 Comparison with LDA. To examine that STM is an appropriate topic modeling algorithm for the task of automating the inductive analysis of mind maps, a comparison between STM and LDA was conducted. LDA was applied to the same datasets with the same values of k, 36, and 34. The package "topicmodels" in R was used to run the LDA algorithm (Hornik & Grün, 2011).

The exact two-step measure used for testing the performance of STM was used to test the performance of LDA against the manual assignment of categories. First, The Chi-square test was used to examine the consistency of the top three topics assigned by LDA to documents with the manual categories of these documents. Second, the cosine similarity between LDA's β_k and manual categories' β_c were computed.

4.6.6 Entrepreneurship Education Application of STM

One of the aims of this research is to enhance EE research with an automated qualitative analysis framework that allows for advanced statistical analysis. Along with automating the inductive analysis of mind maps, STM provides a built-in statistical analysis, linear regression, to estimate the effect of a factor of interest, e.g., academic major, on generated latent topics that represent data-driven entrepreneurial motivations.

This section demonstrates how the automated inductive analysis of mind maps can be utilized in a real EE case. The STM was applied to automate the inductive analysis of mind maps to compare the entrepreneurial motivation of college students belonging to different academic colleges.

One of the reasons for studying the effect of academic majors on the entrepreneurial motivation of college students is to recognize whether EE should be uniformly taught across colleges, e.g., (Duval-Couetil, Reed-Rhoads, & Haghighi, 2012; Franco, Haase, & Lautenschläger, 2010; Kolvereid & Moen, 1997; Matlay, Taatila, & Down, 2012; Pihie & Bagheri, 2011; R. Shinnar, Pruett, & Toney, 2009). Matlay, Taatila, and Down (2012, p. 744) state that "*when teaching entrepreneurship in different academic programs, one cannot help but notice the dissimilarities in the views of their students on the subject*." In the same context, but from a different perspective, Hynes (1996) argues that EE should be promoted among students from non-business majors; thereby, for an active EE intending to help students from different majors, it is essential to understand the differences and similarities among students in terms of entrepreneurial motivations (Maresch et al., 2016).

Many studies to explore entrepreneurial motivation among university students have concentrated on business students (Levenburg, Lane, & Schwarz, 2006). Franco, Haase, and Lautenschläger (2010) report that business administration students were noticeably more entrepreneurially motivated than students from other colleges. On the other hand, Tackey, Perryman, and Connor (1999) find that the highest self-employment rates come from creative arts design courses.

Providing EE researchers with a framework to look into how EE can most effectively improve students' entrepreneurial motivation within the context of academic majors through advanced technology, e.g., STM, is one of the aims of this research. The case problems in this application can be stated as: does a student major affect students' entrepreneurial motivation, specifically are there statistically significant variances between the BUS and CS students' latent topics.

To test the effect of academic major on the entrepreneurial motivation of college students, STM was applied as follows:

- The dataset, "major," includes 1130 motivation documents and 1055 deterrence documents. The metadata for these documents is the academic major, i.e., whether BUS or CS.
- For the two subsets of documents, motivation, and deterrence, searching for *k* and model validation was performed with similar methods presented in section 4.6.4.1. The quality for each value of *k* was measured by "semantic coherence" and "exclusivity".
- The best *k* value in terms of the above metrics was used to run STM.
- A regression analysis was performed to estimate effects of academic major on latent topics found by STM, where documents were the units, a single topic-probability in documents was the dependent variable, and the academic major was the independent variable. The assumptions of linear regression analysis were tested.
- Only the STM topics with statistically significant variance between the two majors were identified, and further analysis was carried out, starting with the interpretation of these topics in the context of entrepreneurial motivations.

Each topic generated by STM includes four distinct types of words; the highest probability, frequency and exclusivity "FREX", Score, and LEFT words, which are described in Section 3.4.2.2. In general, the highest probability words are used as a summarization of topics; nevertheless, using that as a measure tends to prefer words with high frequency overall, but may not hold semantic interest (Roberts, Stewart, & Tingley, 2014). The work of Bischof and Airoldi (2012) exhibited the importance of using exclusivity when summarizing words inside topics, thus, the "FREX" metric was used as a summary for latent topics.

A given topic was defined by its most frequent and exclusive words, "FREX," as well as its most representative documents. *FindThoughts* function from the "*stm*" package was applied to return the most representative documents for each topic (Roberts, Stewart,

& Tingley, 2014). The M-B model was used as a reference to interpret and define the latent topics within the context of EE.

CHAPTER 5

RESEARCH FINDINGS

5.1 Quantitative Content Analysis Results

Table 5.1 shows several summary statistics for the quantitative analysis of mind maps collected in the first dataset, culture. The summary includes a total number of collected mind maps "Total MM," total nodes and tokens, average nodes and tokens per map, average tokens per node, and map's depth. U.S. students wrote more tokens than French students in both the motivation and deterrence mind maps. French students drew more nodes than the U.S. students in the motivation subset only.

The distribution of nodes, tokens, and depth of levels failed the test for normality (Table 5.2). To test for significance of variance, the Mann-Whitney U test was applied as a nonparametric test. Table 5.3 presents that at $p < 0.1$ significance level, the difference in the number of nodes drawn by the U.S. and France students in deterrence maps is statistically significant. The difference in the number of written tokens in the motivation and deterrence maps are statistically significant at $p < 0.01$ significance level.

Table 5.1 Summary Statistics for the Culture Mind Maps by Culture

	Total MM	Total Nodes	Total Tokens	Node per MM	Token per MM	Token per Node	Map Depth
US	21	289	1195	13.76	56.90	4.43	1.77
France	26	298	842	11.46	32.38	2.87	1.61
US	21	278	1335	13.24	63.57	4.96	1.70
France	26	254	675	9.77	25.96	2.80	1.49

Table 5.2 Normality Test - Culture Maps

	Motivation	Deterrence
Node	.000***	.000***
Token	.004***	.000***
Level	.000***	.001***

Table 5.3 P-values for the Variance between the U.S. and French Mind Maps Statistics

	Motivation	Deterrence
Node	.186	.054*
Token	.005**	.000***
Level	.13	.131

Table 5.4 shows summary statistics for mind maps collected in the second datasets, major, sorted by semester. Table 5.5 presents the summary statistics of mind maps drawn by BUS and CS students only. From the demonstrated statistics, similarities between the mind maps of BUS and CS students are seen, except in the number of tokens, where CS students wrote more in both motivation and deterrence maps.

All the distributions of nodes, tokens, and levels, failed test for normality (Table 5.6). Table 5.7 shows that when applying the Mann-Whitney U test, no statistically significant differences between BUS and CS's mind maps have been found.

Table 5.4 Summary Statistics for the Major Mind Maps Sorted by Semester

Semester	Year	Total MM	Total Nodes	Total Tokens	Node per MM	Token per MM	Token per Node	Map Depth
Spring	17	17	265	946	15.59	55.65	3.77	1.77
Fall	17	24	358	1836	14.92	76.5	4.86	1.84
Spring	18	17	190	898	11.18	52.82	4.68	1.59
Fall	18	24	429	1404	17.88	58.5	3.52	1.92
Spring	19	31	386	1561	12.45	50.35	4.1	1.72
Sum		113	1628	6645				
MEAN		22.6	325.6	1329	14.404	58.764	4.186	1.768
Spring	17	17	222	896	13.06	52.71	4.19	1.84
Fall	17	24	308	1651	12.83	68.79	5.44	1.76
Spring	18	17	154	873	9.06	51.35	5.55	1.52
Fall	18	24	447	1606	18.625	66.92	3.86	1.85
Spring	19	31	354	1511	11.42	48.74	4.22	1.67
Sum		113	1485	6537				
MEAN		22.6	297	1307.4	12.999	57.702	4.652	1.728

Table 5.5 Summary Statistics for BUS and CS Mind Maps

	Total MM	Total Nodes	Total Tokens	Node per MM	Token per MM	Token per Node	Map Depth
BUS	38	549	2021	14.45	53.18	3.78	1.79
CS	42	581	2494	13.83	59.38	4.43	1.78
BUS	38	526	2151	13.84	56.61	4.33	1.78
CS	42	529	2505	12.60	59.64	5.14	1.72

Table 5.6 Normality Test - Major Maps

	Motivation	Deterrence
Node	.018**	.000***
Token	.000***	.000***
Level	.000***	.000***

Table 5.7 P-values for The Variance between BUS and CS Mind Maps Statistics

	Motivation	Deterrence
Node	.653	.147
Token	.349	.732
Level	.892	.490

The top two charts in Figure 5.1 show the distribution of nodes over levels in the culture dataset, and the bottom two charts show the distribution of tokens over levels. When asked about deterrence, French students reached three levels, while U.S. students extended their maps to four levels. As seen in Table 5.1, U.S. maps hold more depth with an average of 1.73 levels per map. Level 2 holds more nodes and tokens than any other level, except in the U.S. deterrence maps, where an average of tokens in level 4 exceeded other levels.

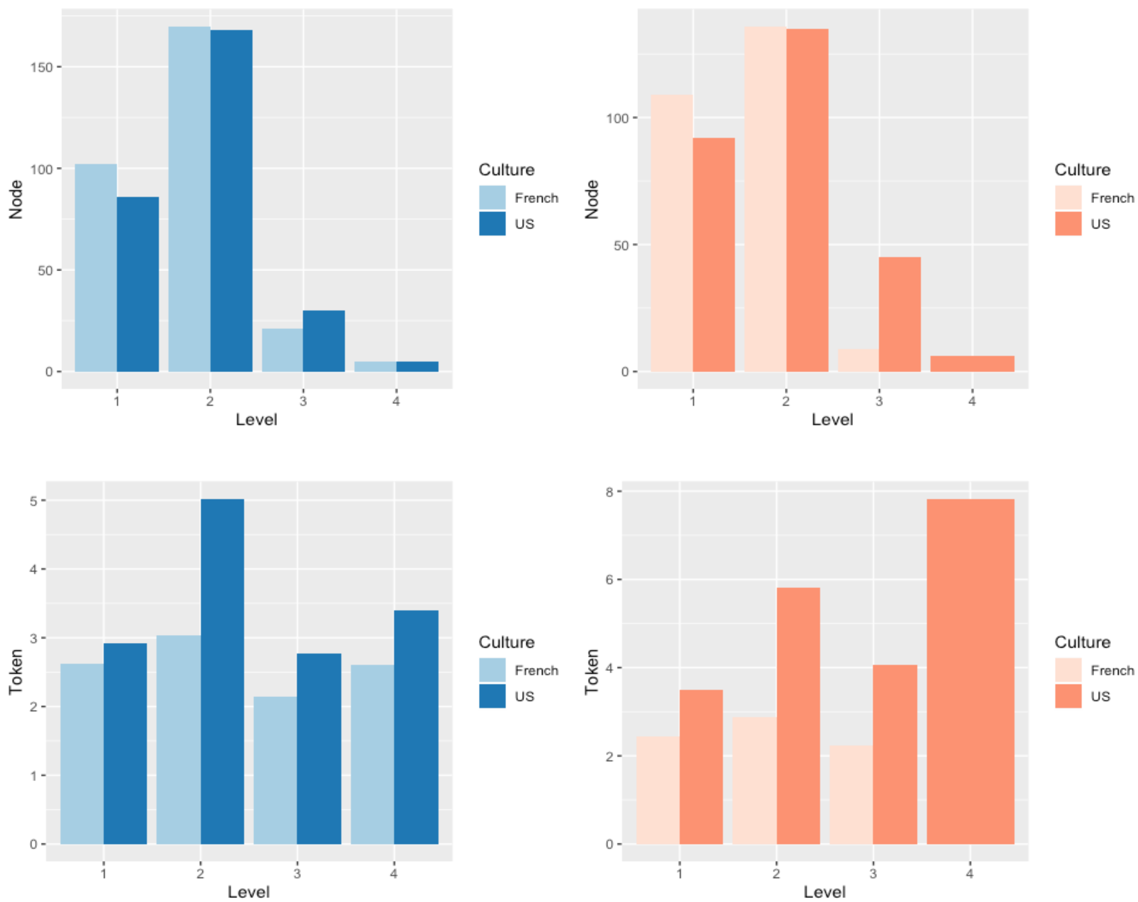

Figure 5.1 Histograms for culture mind maps presenting the number of nodes and tokens per level in motivation, in blue, and deterrence, in orange.

The top two charts in Figure 5.2, the distribution of nodes over levels, show similar patterns between BUS and CS maps. CS students wrote more tokens in levels 1 and 2, while BUS students wrote more tokens in levels 3 and 4 (bottom two charts in Figure 5.2). CS maps reached level 5 in the motivation and deterrence maps, while BUS maps stopped at level 4. BUS maps, however, presented a higher average of depth, as shown in Table 5.5.

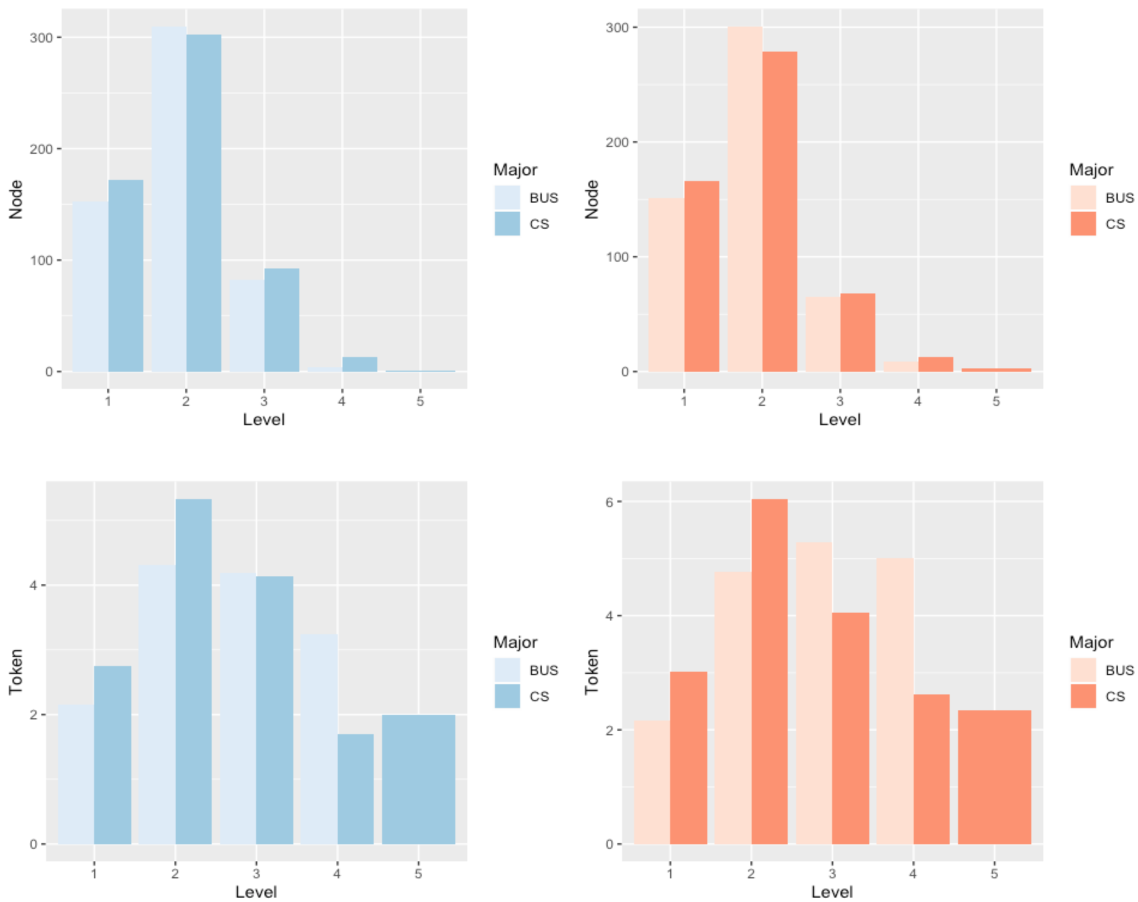

Figure 5.2 Histograms for "major" mind maps showing the number of nodes and tokens per level of a mind map in motivation, in blue, and deterrence, in orange.

5.2 Findings of Experiments

5.2.1 Findings of Experiment Number One

Table 5.8 presents the significance of text classification performance operated on six different feature representations for culture-motivation subset. The p-values were computed from the best models in terms of accuracy across the same feature representation, their accuracies are highlighted in bold fonts. The Regularized LR algorithm has achieved the best accuracy in three different feature representations, in which all were statistically significant at $p < 0.05$ level. Meanwhile, RF performance exceeded that of LR in all the

SVD representations. LR applied to NPMI feature representation has achieved the best accuracy among all algorithms and feature representations, acc = .624.

Table 5.9 exhibits the F-score for the motivation classifiers. The F-score is introduced because the classes in the motivation subset are imbalanced, as shown in Figure 4.8. LR achieved the most significant performance among all classifiers with the TF feature representation, F-score = .588.

Table 5.8 Text Classification Accuracy with Models Significance for Motivation

Feature Rep.	**RF**	**SVM**	**KNN**	**LR**	**P-value**	**Sig.**
TF	0.586	0.553	0.580	**0.620**	0.0220	*
TFIDF	0.617	0.593	0.512	**0.620**	0.0293	*
NPMI	0.610	0.569	0.593	**0.624**	0.0106	*
SVD-50	0.559	**0.593**	0.520	0.554	0.0278	*
SVD-100	**0.554**	0.531	0.503	0.542	0.1084	
SVD-200	**0.582**	0.565	0.514	0.503	0.0420	*
	Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

Table 5.9 Text Classification F-Score for Motivation

Feature Rep.	**RF**	**SVM**	**KNN**	**LR**
TF	0.552	0.487	0.556	**0.588**
TFIDF	0.562	0.551	0.538	**0.566**
NPMI	**0.576**	0.494	0.579	0.568

Table 5.10 shows the significance test of classifiers performance applied to six different feature representations for the culture-deterrence subset. The reported p-values were computed from the best models in terms of accuracy across the same feature representation; the best accuracy for each feature representation is highlighted in bold font.

RF algorithm has achieved the best accuracy in four feature representations, with a significant level at $p < 0.001$. RF achieved the most significant performance among all classifiers with the SVD-100 feature representation, accuracy = .673.

Table 5.10 Text Classification Accuracy with Models Significance for Deterrence

Features Rep.	**RF**	**SVM**	**KNN**	**LR**	**P-value**	**Sig.**
TF	0.585	0.585	0.528	**0.604**	0.0000	***
TFIDF	0.604	0.611	0.570	**0.657**	0.0000	***
NPMI	**0.630**	0.630	0.551	0.611	0.0000	***
SVD-50	**0.610**	0.597	0.572	**0.610**	0.0000	***
SVD-100	**0.673**	0.623	0.635	0.616	0.0000	***
SVD-200	**0.610**	0.597	0.522	0.591	0.0000	***
	Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

The above results of classifiers' accuracy and significance levels have shown that text classification can be applied to automate deductive analysis of mind maps. Because the human analysis was used as the baseline for these measures, it was concluded that text classification models' performance is similar to that of human analysis, thus, the hypothesis H_{1a} was supported.

For testing the significance of mind map topology under the two methods of feature extraction, classifiers A (not including topological features) and B (including the topological features) have used the same hyperparameters when predicting the test set. This allowed for better comparison since the only difference between the two classifications was only the features of mind map topology (Salzberg, 1997).

Table 5.11 shows the Wilcoxon test results for testing the effect of inserting mind map topology as features on the classification task under the two methods. The p-value of the two tests was not significant, thus, both hypotheses H_{2a} and H_{2b} were not supported. The paired t-test has reported equivalent results for the comparisons.

The MDG scores representing information gain given by the RF algorithm indicated that the map topological features under the two methods were important. Table 5.12 presents the results of the feature importance computed by MDG for the general-link method. Two features out of the four, representing the branch and nodes in level 2, have gained higher MDG than most of the data's features. Under the specific-link method, Table 5.13 shows the feature importance measured by MDG, and also, the added features scored higher than any other features in the data.

Even though the improvement achieved by using mind map topology as features under the two methods was not statistically significant, the extracted features have gained higher MDG scores than any of the remaining features, which is an indication of their importance for the classification task.

Table 5.11 The Wilcoxon Signed-Ranks Test Results for Comparison between Classifier A and B Under the Two Methods

Method	P-value
General-link (1st)	0.1763
Specific-link (2nd)	0.3394

Table 5.12 Feature Importance by Random Forest for Mind Map Topology Under Method One (General-link)

Feature	Mean Decrease Gini
L1	10.60
L2	8.12
L3	2.69
fear	4.74
failure	6.65

Table 5.13 Feature Importance by Random Forest for Mind Map Topology Under Method Two (Specific-link)

Feature	Mean Decrease Gini
map	29.18
branch	11.09
fear	7.8
failur	6.18
financi	1.98
difficulti	1.64

It is apparent that across all tasks, either RF or LR performs best among the models, and as many studies have observed, no single classification algorithm is the best for all cases and datasets (Salzberg, 1997). It is also imperative to note that the model that finds it hard to learn from the training set produced better performance estimates for classifier effectiveness; this agreed with the findings of Domingos (2012). A model is anticipated to poorly generalize the testing set when overfitted on the training set and the possibility of generating worse accuracies than other models with better regularization.

For text classification tasks, textual feature representation is a crucial element that can upgrade or downgrade the overall performance; the experiment included four different representations. As the results showed, NPMI has produced the best average accuracy for all the four algorithms with .599, and TF-IDF came second with .586 for the motivation subset. SVD-100 obtained the best average accuracy with .637 for the deterrence subset. The implementation of various feature representations was essential to the intention of conducting extensive experimentation of text classification.

5.2.1.1 Findings of Application – Culture Case Problems. Table 5.14 shows that three of the weights are normally distributed. A regression analysis was performed to test the case problems involving these weights. For the weights of stage four, s_4, in the deterrence set, Mann-Whitney U test was applied, and a regression analysis was also performed to explain the variance of s_4, the p-value for the regression is considered to be significant only with an alpha level < 0.05, as suggested by Kutner et al. (2005).

Table 5.14 Shapiro Normality Test for Stages' Weights

Data	Weight	P-value	Result
Motivation	s_1 “Innovation”	0.3179	Normal
	s_4 “Growth”	0.3767	Normal
Deterrence	s_1 “Innovation”	0.1157	Normal
	s_4 “Growth”	0.0072	Not Normal

Table 5.15 shows that when asked about motivation, U.S. students are more likely to prefer the innovation stage than French students, with a significance level of p < 0.001. On the other hand, U.S. students are less likely to prefer the growth stage than French students with a significance level of p < 0.01.

Table 5.16 shows that when asked about deterrence, U.S. students are less likely to prefer the innovation stage than French students with a significance level of p < 0.01; and they are more likely to prefer the growth stage than French students with a significance level of $p < 0.01$. For the Growth weight in the deterrence documents, the Mann-Whitney U test reported a significant p-value, 0.0083; the Growth weight's rank among the U.S. and French students is not equal to zero.

Table 5.15 The Regression Analysis Results for Motivation

Motivation	**Culture**	**Estimate**	**Std. Error**	**t value**	**Sig.**	
Innovation	U.S.	0.1248	0.0294	4.2430	0.0001	***
Growth	U.S.	-0.1448	0.0475	-3.0490	0.0038	**
	Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

Table 5.16 The Regression Analysis Results for Deterrence

Deterrence	**Culture**	**Estimate**	**Std. Error**	**t value**	**Sig.**	
Innovation	U.S.	-0.0675	0.0233	-2.8970	0.0058	**
Growth	U.S.	0.1359	0.0442	3.0740	0.0036	**
	Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

Therefore, all the following case problems were supported:

- *Case Problem 1$_a$*: $s_1|_{\text{US, motivation}} > s_1|_{\text{France, motivation}}$
- *Case Problem 1$_b$*: $s_1|_{\text{France, deterrence}} > s_1|_{\text{US, deterrence}}$
- *Case Problem 1$_c$*: $s_4|_{\text{France, motivation}} > s_4|_{\text{US, motivation}}$
- *Case Problem 1$_d$*: $s_4|_{\text{US, deterrence}} > s_4|_{\text{France, deterrence}}$

5.2.2 Findings of Experiment Number Two

The Chi-square test of independence was used to test the association between the STM latent topics and the manual categories, attributes of the M-B model, assigned to the same document. The test was performed, first, on the manual categories and the dominant topics of STM, i.e., topics having the highest probability inside documents. Then, between the manual categories and the second top topics, 2nd Topic, and finally, between the manual categories and the third top topics, 3rd Topic.

Table 5.17 shows the results of the test for the motivation set. Table 5.18 presents the results for the deterrence set. Statistically significant associations between the manual categories assigned to documents and with each of the top three topics generated by STM for that document were inferred. The STM assignment of latent topics was illustrated to be statistically consistent with the manual assignment of categories. The results show that the dominant topics, in both the motivation and deterrence subsets, are associated with the manual categories with a significance level of $p < 0.001$.

Table 5.17 Results of Testing Association between STM and Manual Analysis for Motivation Set

	χ^2	P-value	Sig.
Dominant Topics	299.37	.0000	***
2nd Topics	220.67	.0000	***
3rd Topics	360.66	.0000	***

Significance: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 5.18 Results of Testing Consistency between STM and Manual Analysis for Deterrence

	χ^2	P-value	
Dominant Topics	252.35	.0000	***
2nd Topics	296.95	.0000	***
3rd Topics	264.35	.0000	***

Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 5.19 shows the results of cosine similarity between each of the top three topics' β_k and the manual categories' β_c across documents. The similarity between β_k of the dominant topics, denoted by β_{t1}, and $\boldsymbol{\beta_c}$ reached .213 in the motivation documents and .281 in the deterrence documents. The similarity between β_k and β_c supports the results produced by the Chi-square test. Besides the statistically significant consistency between human analysis and STM, also the probabilities of the words over them are similar.

The hypothesis of H_{1b} "the performance of automating the inductive analysis of mind maps collected in EE research with STM is similar to human analysis," was supported by the two-step measure. STM was effectively capable of automating the inductive analysis of mind maps. STM has discovered latent topics consistent with the categories assigned by human analysis, and it was also effective in generating word probability over topics that are similar to the probability of words over manual categories.

Table 5.19 Cosine Similarities between the STM Top 3 Topics' β_k and Manual Assignment's β_c

	SIM $(\boldsymbol{\beta_{t1}}, \boldsymbol{\beta_c})$	**SIM** $(\boldsymbol{\beta_{t2}}, \boldsymbol{\beta_c})$	**SIM** $(\boldsymbol{\beta_{t3}}, \boldsymbol{\beta_c})$
Motivation	0.213	0.131	0.124
Deterrence	0.281	0.149	0.115

The benefit of treating nodes as the unit of analysis, i.e., documents, was proven to improve the STM topic quality by assigning nodes to as close as a single topic, which is comparable to that of human analysis. Table 5.20 shows the exact binomial test for the dominant topic probability inside all documents with a 95% confidence interval. The sample estimate was statistically more significant than the null estimate of $\hat{p}$ = .5.

STM has assigned nodes proportionally into as close as a single topic, with an average estimate of $\mathring{p}$ = .55. This implies that the dominant topics probability was more prevalent than the sum of all remaining topics probability, and therefore, the hypothesis H_{2c} was supported.

To illustrate the results, Table 5.21 shows the average of the STM's top five topic probability inside nodes. The table exhibits the significant difference between the probability of the dominant topic and the second top topic inside one node. The dominant topic was about 44% more prevalence in one node than the nearest topic. These results have also supported the assumption of examining only the top three topics inside documents against the responding manual categories, where the top three topics represented more than 65% of total θ_d, i.e., probability distribution of topics over documents.

Table 5.20 The Results of the Exact Binomial Test

Null probability	0.5
Alternative H1	True p is greater than 0.5
P-value	0.0001 ***
Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1	

Table 5.21 Average of the Top Five Topics Probability in All Documents

Topics	1st	2nd	3rd	4th	5th
Probability	0.554	0.105	0.040	0.035	0.012

For the comparison between STM and LDA, Table 5.22 shows the results of the Chi-square test of independence applied to the LDA latent topics and the manual categories, the results were not statistically significant. The LDA latent topics have no consistency with the manual categories, meaning that LDA has generated latent topics that were not consistent with the assignment of manual categories. This contradicts the findings of the STM latent topics, where consistency was statistically significant to human analysis. LDA could also not assign nodes proportionally into as close as a single topic. The average of dominating topics probability was $\mathring{p}$ = .28 per node, which is less than the null probability of $\hat{p}$ = .5.

Table 5.22 Chi-square Test Results for Manual Categories and LDA Latent Topics

Motivation	χ^2	P-value
1st Topics	773.2823	0.1618
2nd Topics	755.0340	0.3644
3rd Topics	553.0582	0.7541
Deterrence		
1st Topics	802.3548	0.4015
2nd Topics	851.3837	0.1365
3rd Topics	535.8067	0.7678

The computation of cosine similarity between the β_c of manual categories and the β_k of LDA is presented in Table 5.23. The similarities are not as high as the cosine similarities achieved by STM, see Table 5.19.

The β_k of LDA was not as meaningful as the β_k of STM. Table 5.24 shows a sample of β_k for five LDA topics, where words hold almost the exact probability of importance to these topics. On the other hand, the STM words probability over topics is apparent and easy to distinguish. For example, Figure 5.3 presents the importance of words to topic number 1 and topic number 2 of STM. This information indicates that STM's β_k identifies unique words for each latent topic, permitting better word assignments into topics, unlike LDA results. For all these reasons, the hypothesis H_{1c} was supported and STM outperforms LDA when both used to automate inductive analysis of mind maps.

Table 5.23 Cosine Similarities between the LDA Top 3 Topics' β_k and Manual Assignment's β_c

	SIM ($\boldsymbol{\beta_{t1}}, \boldsymbol{\beta_c}$)	**SIM ($\boldsymbol{\beta_{t2}}, \boldsymbol{\beta_c}$)**	**SIM ($\boldsymbol{\beta_{t3}}, \boldsymbol{\beta_c}$)**
Motivation	0.064	0.061	0.060
Deterrence	0.063	0.086	0.092

Table 5.24 LDA Beta of a Sample of Words for Five Topics

Word	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
boss	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
break	0.0010672	0.0012392	0.0010672	0.0009930	0.0134639
build	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
busi	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
butt	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
buy	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
can	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
capabl	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
capit	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240
career	0.0010672	0.0012392	0.0010672	0.0009930	0.0012240

The STM's outperformance of LDA was mainly due to the STM ability to incorporate metadata of interest, such as culture. The STM discovery of latent topics does not solely depend on texts inside documents, but also on the affiliation of that text, e.g., U.S. or France. STM recognizes the covariate's effect when generating latent topics, which is a distinguishing feature of STM among all topic modeling algorithms. STM was also proven to handle short texts better than LDA, the word-topic probability of STM, β_k, produced more meaningful values than LDA's β_k.

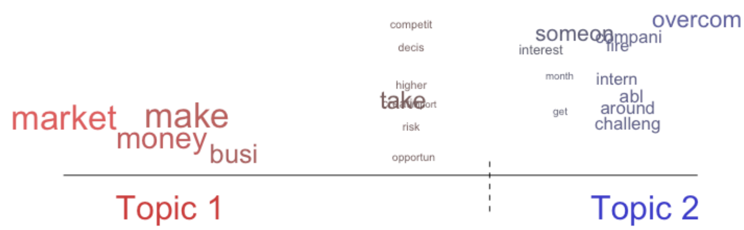

Figure 5.3 The words importance in topic 1 and 2 of the STM. For example, the word "market" is very important to topic one, while the word "overcome" is important to topic two.

5.2.2.1 Findings of Application – Major Case Problems. In this application, the dataset "major" was used. For motivation: STM had 47 topics, 1130 documents, and a 746-word dictionary. For deterrence: STM had 43 topics, 1055 documents, and an 812-word dictionary. For estimating the effects of academic major on latent topics, there were 47 regressions for the motivation subset and 43 for the deterrence subset, one for each latent topic.

Table 5.25 shows the STM latent topics in the motivation documents significantly affected by the major academic covariance. The reference level for the analysis was BUS, and major2 reported in the table refers to CS. The remaining latent topics, 35 topics, were not significantly impacted by the academic major; they were shared between BUS and CS students.

Table 5.26 exhibits the topics most frequent and exclusive words "FREX." Figure 5.4 shows an example of most representative documents for topic one, "Social – Personal Values 1". The topics' most frequent and exclusive words "FREX" and representative documents have been both used to define the STM latent topics as entrepreneurial motivations. Figure 5.5 shows the significant coefficients of academic majors with a 95% confidence interval, where the x-axis represents CS coefficients.

Table 5.25 The Results of Estimating the Effect of Academic Majors on STM Latent Topics - Motivation

Topic			**Estimate**	**Std. Err**	**t value**	**Sig.**	
Social – PV 1	1	(Intercept)	0.0300	0.0041	7.2521	0.0000	
		major2	-0.0130	0.0058	-2.2425	**0.0251**	*
Social – PV 2	3	(Intercept)	0.0320	0.0047	6.7881	0.0000	
		major2	-0.0159	0.0067	-2.3718	**0.0179**	*
Environment	4	(Intercept)	0.0063	0.0023	2.7836	0.0055	
		major2	0.0070	0.0033	2.1146	**0.0347**	*
Leadership 1	14	(Intercept)	0.0109	0.0029	3.7079	0.0002	
		major2	0.0104	0.0048	2.1527	**0.0316**	*
Passion	15	(Intercept)	0.0395	0.0049	7.9970	0.0000	
		major2	-0.0212	0.0059	-3.6112	**0.0003**	***
Leadership 2	19	(Intercept)	0.0148	0.0038	3.9395	0.0001	
		major2	0.0120	0.0055	2.1856	**0.0291**	*
Team	22	(Intercept)	0.0168	0.0042	4.0496	0.0001	
		major2	0.0183	0.0060	3.0463	**0.0024**	**
Leadership 3	23	(Intercept)	0.0235	0.0031	7.6622	0.0000	
		major2	0.0155	0.0047	3.3172	**0.0009**	***
Person-Values	25	(Intercept)	0.0373	0.0046	8.1501	0.0000	
		major2	-0.0183	0.0062	-2.9477	**0.0033**	**
Strategy	36	(Intercept)	0.0149	0.0038	3.9694	0.0001	
		major2	0.0191	0.0052	3.7102	**0.0002**	***
Sustainability	40	(Intercept)	0.0064	0.0025	2.5567	0.0107	
		major2	0.0085	0.0038	2.2577	**0.0242**	*
Opportunity	42	(Intercept)	0.0338	0.0044	7.7706	0.0000	
		major2	-0.0167	0.0054	-3.0896	**0.0021**	**
		Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

Table 5.26 The Significant STM Latent Topics FREX words - Motivation

Topic	EM							
1	**Social – PV 1**	dream	inspir	proud	other	pursu	desire	serv
3	**Social – PV 2**	give	back	one	class	competit	poor	voic
4	**Environment**	extens	field	educ	sort	experi	level	offici
14	**Leadership 1**	possibl	everi	earn	aspect	reput	smart	strong
15	**Passion - PV**	fame	pride	invent	disabl	expans	recognit	reinvent
19	**Leadership 2**	improv	industri	process	entertain	well	struggl	continu
22	**Team**	like	peopl	mani	play	relax	hire	alway
23	**Leadership 3**	work	passion	futur	balanc	els	done	someon
25	**Person-Values**	live	free	debt	thing	comfort	enjoy	posit
36	**Strategy**	fund	follow	will	ventur	mentor	model	role
40	**Sustainability**	sustain	fresh	produc	food	local	good	tri
42	**Opportunity**	opportun	explor	creativ	art	love	learn	share

The students were asked about "things that motivate them to create a new venture"; thus, the following results were inferred from the regression analysis:

- CS students were less likely to be inclined to Social Personal Values, topic 1 and 3, as entrepreneurial motivations than BUS students, with significance level at $p < 0.05$.
- CS students were less likely to be inclined to Passion, topic 15, as an entrepreneurial motivation than BUS students, with significance level at $p < 0.001$.
- CS students were less likely to be inclined to Opportunity, topic 42, as an entrepreneurial motivation than BUS students, with significance level at $p < 0.01$.

On the other hand,

- CS students were more likely to be inclined to Environment of implementation and Sustainability, topic 4 and 40, as entrepreneurial motivations than BUS students, with significance level at $p < 0.05$.
- CS students were more likely to be inclined to Leadership 1-2, in topics 14 and 19, as entrepreneurial motivations than BUS students, with significance level at $p < 0.05$.

- CS students were more likely to be inclined to Leadership 3, in topic 23, as an entrepreneurial motivation than BUS students, with significance level at $p < 0.001$.

- CS students were more likely to be inclined to Team, topic 22, as an entrepreneurial motivation than BUS students, with significance level at $p < 0.01$.

- CS students were more likely to be inclined to Strategy, topic 36, as an entrepreneurial motivation than BUS students, with significance level at $p < 0.001$.

Representative Documents for Topic 1:

"inspiring others to pursue their dreams - serving others brings me joy - desire for material possessions - desire to do well by others - becoming a leader for others - enables me to dream big - inspire others to be proud of myself and my actions - pursue my dreams - share creativity with others to inspire"

Figure 5.4 Sample of most representative documents for latent topic 1 "Social Personal Values."

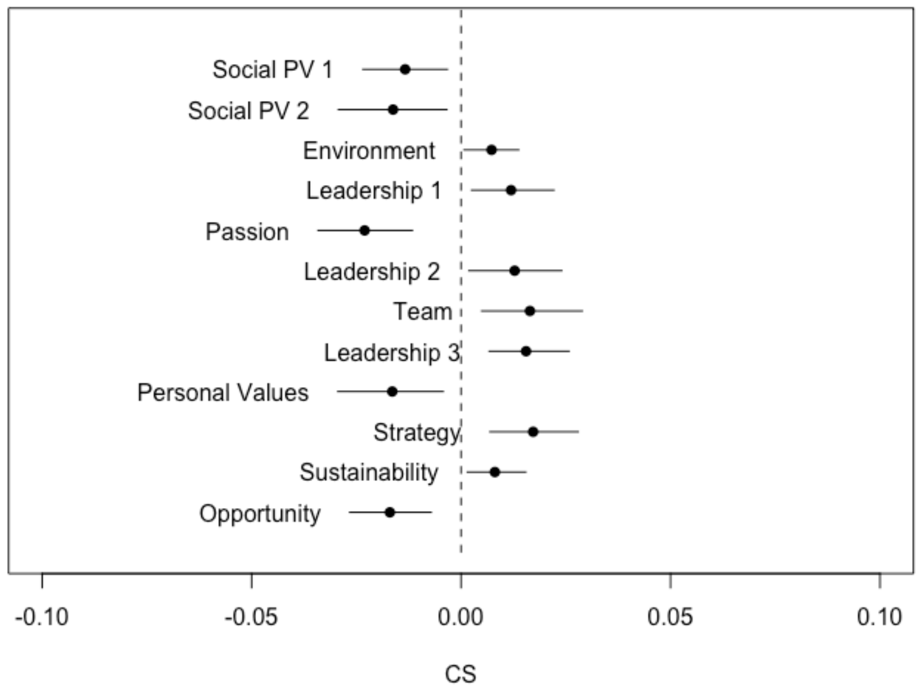

Figure 5.5 Estimating the effect plot of CS on the STM latent topics "motivation." It is a positive association for topics' coefficients in the positive side and a negative association for topics' coefficients on the negative side.

All the remaining STM latent topics in the motivation documents were not significantly affected by the academic majors. BUS and CS students had shared topics that included Family, Flexibility, n-Achievement, Investment, Resources, Marketing, Learning, Problem Solving, and Innovation as entrepreneurial motivations.

Table 5.27 shows the statistically significant effect of academic majors on the STM latent topics for the deterrence documents. Table 5.28 exhibits the most frequent and exclusive words of these topics, and Figure 5.6 demonstrates an example of most representative documents for a given topic, which have guided the interpretation of latent topics extracted from the deterrence documents. Further, Figure 5.7 shows the confidence interval of 95% of the significance among these topics, where the x-axis represents CS coefficients.

Table 5.27 The Results of Estimating the Effect of Academic Majors on STM Latent Topics - Deterrence

Topic			Estimate	Std. Err	t value	Sig.	
Ambiguity Tolerance 1	6	(Intercept)	0.0249	0.0039	6.3728	0.0000	
		major2	0.0108	0.0063	1.7207	**0.0597**	.
Leadership	12	(Intercept)	0.0117	0.0037	3.1675	0.0015	
		major2	0.0106	0.0058	1.8270	**0.0313**	.
Risk Taking	19	(Intercept)	0.0178	0.0043	4.1308	0.0000	
		major2	0.0180	0.0071	2.5206	**0.0175**	*
Competition	25	(Intercept)	0.0349	0.0054	6.4510	0.0000	
		major2	-0.0233	0.0067	-3.4879	**0.0006**	***
Ambiguity Tolerance 2	28	(Intercept)	0.0097	0.0042	2.3190	0.0205	
		major2	0.0272	0.0065	4.1631	**0.0000**	***
Management	31	(Intercept)	0.0262	0.0042	6.2309	0.0000	
		major2	-0.0135	0.0054	-2.5240	**0.0157**	*
Resources	42	(Intercept)	0.0318	0.0049	6.4659	0.0000	
		major2	-0.0136	0.0068	-2.0162	**0.0266**	*
		Significance: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

Table 5.28 The Significant STM Latent Topics FREX words - Deterrence

Topic	EM						
6	**Ambiguity Tolerance 1**	consum	focus	success	grow	lost	broke
12	**Leader**	may	negat	partner	even	amount	possibl
19	**Risk-Taking**	person	financ	issu	capit	health	sacrif
25	**Competition**	custom	boss	chang	plan	complet	expens
28	**Ambiguity Tolerance 2**	mind	like	revenu	mentor	option	less
31	**Management**	alway	disadvantag	deliv	first	alreadi	low
42	**Resources**	financi	stabil	feel	industri	guess	second

In the deterrence documents, students answered for “things that deter them from creating a new venture,” Hence, the following results were inferred:

- CS students were more likely to be deterred by Ambiguity Tolerance 1, topic 6, than BUS students, with significance level at $p < 0.1$.
- CS students were more likely to be deterred by Leadership, topic 12, than BUS students, with significance level at $p < 0.1$.
- CS students were more likely to be deterred by Risk-Taking, topic 19, than BUS students, with significance level at $p < 0.05$.
- CS students were more likely to be deterred by Ambiguity Tolerance 2, topic 28, than BUS students, with significance level at $p < 0.001$.

On the other hand,

- CS students were less likely to be deterred by Competition, topic 25, than business students, with significance level at $p < 0.001$.
- CS students were less likely to be deterred by Management, topic 31, than BUS students, with significance level at $p < 0.05$.
- CS students were less likely to be deterred by Resources, topic 42, than BUS students, with significance level at $p < 0.05$.

Representative Documents Topic 6:

“the EPA does not approve of my growing methods - lost respect and identity - doubting my abilities to be successful - failure might send them back to home countries - plan not guaranteed to be successful - no guarantees of success - no guarantee success - what are the chances of success”

Figure 5.6 Sample of most representative documents for latent topic 6 " Ambiguity Tolerance 1."

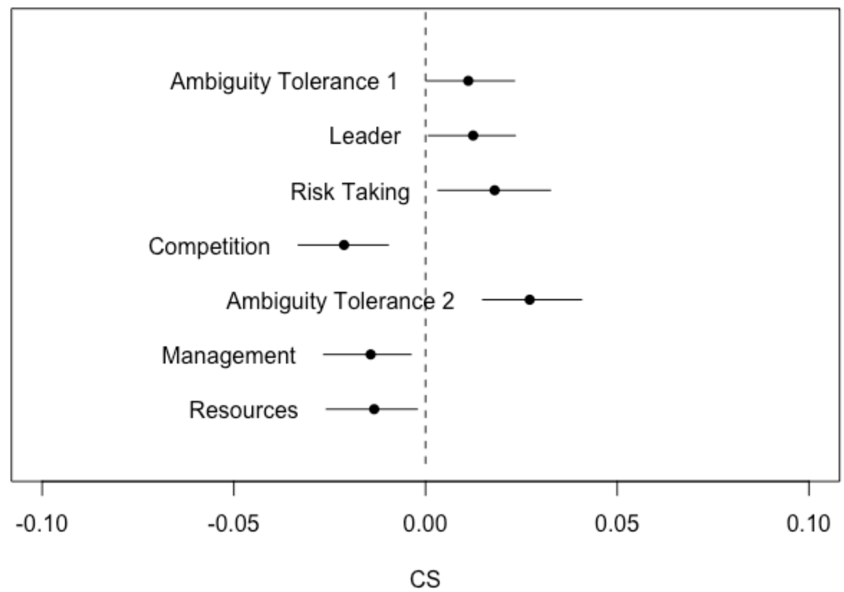

Figure 5.7 Plot of estimating the effect of computer science major on the STM latent topics “deterrence.” It is a positive association for topics’ coefficients in the positive side and a negative association for topics’ coefficients on the negative side.

5.3 Conclusion

Table 5.29 exhibits the summary of hypotheses and their testing results. The process automation hypotheses introduced in this research have been statistically supported. The text classification models have successfully automated the deductive analysis, and STM has successfully automated the inductive analysis of mind maps. The performance of text classification models was statistically significant compared to the human deductive analysis, and the performance of STM was statistically significant compared to robust human analysis.

Using mind map topology as features have improved the accuracy and F-score metrics, only under the second method, for the classification task despite failing to present a statistically significant difference. The feature selection implemented by RF “MDG” also signified the importance of two out of the four mind map topology features under method number one. The other two did not gain as high score because they represented nodes' positions in levels three and four, which were not as representative as level one and two features. Figure 5.1 showed that most nodes inside mind maps are in levels one and two. Under method two, both features obtained a high MDG score by RF, especially, the feature "map" representing to which mind map a node belongs.

In the second experiment, treating nodes as documents has improved the quality of STM topics by allowing nodes to be dominated by one latent topic. The domination of one topic per node decreases the ambiguity of semantic meaning that might occur if a document is fragmented over several topics. The domination of the topic per node indicates the efficiency of STM to generate findings that are aligned with the human analysis of similar data (Roberts et al., 2014; Tattersall, Watts, & Vernon, 2007; Bandera et al., 2018b; Wheeldon & Ahlberg, 2017).

The comparison between STM and LDA indicated that for automating mind maps analysis, STM outperformed LDA. Whereas comparing the human analysis to the STM findings were statistically significant, the LDA findings failed the same statistical test procedure. LDA produces poor assignment of latent topics, as well as fragmented word-topic probabilities that were hardly meaningful. LDA also has only reached an average of $\mathring{p} = .28$ as the topic probability per node, while the STM latent topics showed a dominance of one topic probability per node at $\mathring{p} = .55$.

The two EE applications have demonstrated the applicability and feasibility of using the automation framework presented in this dissertation into EE problems. The implementation of text classification and STM to automate the qualitative content analysis of mind maps collected in EE studies are novel.

Table 5.29 The Summary of Hypotheses

Hypotheses	**Supported**
H_{1a}: The performance of automating the deductive analysis of mind maps with classification models is similar to that of human deductive analysis	Yes
H_{1b}: The performance of automating the inductive analysis of mind maps with Structural Topic Model is similar to that of human inductive analysis	Yes
H_{1c}: Structural Topic Model outperforms Latent Dirichlet Allocation for automating the inductive analysis of mind maps	Yes
H_{2a}: Using topology of mind map "General-link" as features improves classification performance when automating the deductive analysis of mind maps	No
H_{2b}: Using topology of mind map "Specific-link" as features improve classification performance when automating the deductive analysis of mind maps	No
H_{2c}: Treating nodes as the unit of analysis improves automating the inductive analysis of STM by assigning nodes to a single "dominant" topic that exceeds the null probability of $\hat{p} = .5$	Yes

CHAPTER 6

DISCUSSION, CONCLUSION, AND SUGGESTIONS FOR FUTURE RESEARCH

6.1 Introduction

The role that qualitative methods can play to enhance EE research can be significant (Gartner & Birley, 2002). Qualitative methods allow participants to articulate their answers and experiences in a way that quantitative methods cannot allow (Rahman, 2017). The importance of applying qualitative methods in EE research is due to the comprehension of context that can be gained by them (Gartner & Birley, 2002). Despite all that, EE research suffers a lack of using qualitative methods and one of the main reasons for that shortage is the difficulty of qualitative analysis (Lorz, Mueller, & Volery, 2013). The manual analysis of qualitative methods is labor-intensive and time-consuming and thereby not scalable to extensive studies involving many students and regions (Graneheim, Lindgren, & Lundman, 2017).

The automation framework introduced in this dissertation integrates machine learning algorithms and QCA approaches to automate the analysis of mind maps used in EE assessment. The automation framework provides EE researchers with automated qualitative research methods that include utilizing mind maps as a data collection tool and applying NLP techniques to automate their QCA.

QCA is suitable to analyze mind maps collected in EE studies because it is a flexible qualitative methodology, it describes data that requires some degree of interpretation, and offers distinct approaches to analyze data (White & Marsh, 2006; Schreier, 2012; Mayring, 2004). The deductive approach was automated with text classification models and the inductive approach was automated with STM.

Implementing machine learning and statistical NLP techniques into the mind maps analysis paves the way for entrepreneurship educators to interpret and understand students' input at a larger scale. Besides shortening the labor, the time required for analysis, and solving for consistency problems, the advanced text analytics can extract and discover findings and knowledge that human analysis does not attain, e.g., (Mankad et al., 2016; Popping, 2015; Reich et al., 2014). The use of STM to automate inductive qualitative analysis allows EE researchers to contextually investigate data and then link findings to theories that can explain such phenomena.

This study also contributes to increasing the sample size in EE qualitative research to mitigate the generalizability and comparability issues (Rahman, 2017). Using this framework makes a large sample size manageable and not affected by time-consuming nor labor-intensive issues. The analysis of mind maps as a data tool is faster than standard qualitative tools (Burgess-Allen & Owen-Smith, 2010), and with automating their analysis, the sample size of a qualitative study can be compatible to a quantitative study.

The automation framework contributes to EE best practices by assisting researchers to qualitatively examine the impact of EE on a large scale that involves students from different societies, cultures, politics, and economies background. The automation framework can stimulate the use of qualitative methods in the field of EE because it allows for collecting extensive qualitative data (mind maps), selecting which QCA approach to implement (deductive or inductive), automating the analysis with modern machine learning techniques (text classification models and STM), and validating the results. The automation framework also permits direct statistical testing for qualitative data in EE. The two experiments have demonstrated how the automation framework can be applied to EE problems, and conducting textual statistical inference.

6.2 Discussion

The main research questions in this dissertation are focused on the process automation for the qualitative content analysis of mind maps and the exploitation of map topology in the process. Text classification algorithms were hypothesized to automate the deductive analysis of mind maps. Deductive analysis requires a reference model that guides the coding of text, which the mapping function in classification algorithms analogously carries out (Egami et al., 2018; Scharkow, 2013).

Experiment number one has laid out the implementation and testing of automating deductive analysis with text classification models, and the experiment was performed on mind maps collected in an EE setting. The results of testing the hypothesis were statistically significant; the classification models have successfully automated mind maps analysis with a performance similar to that of human analysis. The human analysis, which manually classified documents into five classes, was used as the baseline for performance testing. LR and RF models achieved statistically significant accuracy and F-score when predicting mind map data from a test set. The classification models could automate the deductive analysis of mind maps. The experiment of applying classification models to analyze mind maps is novel.

The inductive content analysis of mind maps has been automated by applying STM, an unsupervised machine learning technique. STM was hypothesized to accomplish this task. STM is conceptually similar to the inductive approach in that both extract topics and categories directly from data without using prior knowledge nor a reference to guide textual coding (Baumer et al., 2017). The assignments of latent topics by STM was significantly consistent with human analysis, which manually labeled documents into one of the forty-eight attributes from the M-B model.

The distribution of words over STM topics (β_k) and manual categories (β_c), as probability indicating word importance, were also correlated. Despite producing word probability over topics for all words in the data, when only the top 88- and 73-words probability in STM's β_k were compared to the manual's β_c, which includes different sets of words probability, the cosine similarity between the two achieved .21 and .28 in the motivation and deterrence documents, respectively.

The experimentation is the first of its kind in EE research, where a topic modeling algorithm has been applied to analyze students' data to measure their entrepreneurial motivations. It is also a novel experiment for analyzing mind maps with STM.

The comparison between STM and LDA has validated the hypothesis of using STM as a topic modeling algorithm to automate the inductive analysis of mind maps. The LDA results are not as statistically significant as those of STM. Whereas the LDA latent topics were not persistent with the manual assignment of categories and the probability of words within the LDA's topics had a low cosine similarity to the manual probability of words, the STM latent topics were significantly consistent with the manual assignment, and the cosine similarity between the STM and manual's words probability reached .28.

STM outperformed LDA because of its ability to incorporate covariates of interest, such as major. In STM, the discovery of latent topics does not solely depend on texts inside mind maps, but also on information associated with that text, e.g., BUS or CS. STM recognizes the covariate's influence when generating latent topics, which is a distinctive quality of STM among all topic modeling algorithms. STM was also proven to handle short texts better than LDA.

This dissertation also hypothesized that the exploitation of mind map topology could improve process automation's performance. The first experiment tested two methods for extracting mind map topology features and including them with textual features. When comparing the two classifiers' performance, "A" which did not involve mind map topology and "B" which did involve them, mind map topology improved classifier B's performance in terms of accuracy and F-score under the second method (specific-link), but not under the first one (general-link). The variance between the classifiers' performances under both methods, on the contrary, failed to be statistically significant. This insignificance of improvement might be due to using only 12 datasets to compare the classifiers (Table B.1).

The MDG score produced by RF ranked mind map topological features as the most important within the data, and they scored higher than any other features in terms of information gain. The mind map topology showed potential enhancement for automating the deductive analysis, especially the extracted features under the second method (Table B.1). The specific-link method has shown better performance than the general-link method since it allowed the classifier to recognize nodes from different collections. Under the general-link method, the classifier treats all nodes as if they are in one collection, i.e., one large mind map.

On the other hand, the use of mind map topology in experiment number two has improved the process automation of inductive analysis. Treating nodes as documents, i.e., the unit of analysis, allowed STM to assign a node to one topic with an average of $\mathring{p}$ = .54. The second-largest topics only held $\mathring{p}$ = 0.105. One of the advantages of using mind maps as a collection tool is facilitating qualitative analysis by providing already-segmented ideas and concepts (Wheeldon & Ahlberg, 2017; Wheeldon & Faubert, 2009). The LDA was not capable of generating dominant topics, where the model's highest topic probability inside a node was $\mathring{p}$ = .28.

The deterrence subsets in both the culture and major datasets have helped text classification and STM algorithms to produce better results compared to the motivation subsets. Text classification models have achieved .67 accuracy, whereas the highest model in the motivation subset reached .62 accuracy. In STM, the Chi-square results were higher for the deterrence dataset than the motivation subset.

The first EE application under experiment one demonstrated how text classification of mind maps could measure entrepreneurial motivations. The five pre-defined classes, taken from the M-B model, have been used to statistically test the variance of college students' entrepreneurial motivations belonging to two cultures, U.S. and France. Regression analysis was used as statistical testing for the case problems of application one.

In the second EE application, under experiment two, STM has discovered latent topics within students' mind maps and incorporated their academic major as a covariant. The STM latent topics were meaningful as entrepreneurial motivations. The analysis of estimating the effect of academic majors on latent topics reported statistically significant associations between several topics and documents' metadata, BUS and CS. These latent topics were defined within the context of EE by their most frequent and exclusive words "FREX" and representative documents.

6.3 Conclusion

The automation framework is novel and designed to automate the qualitative content analysis of mind maps. It enables EE researchers to use mind maps as a data collection tool, select a QCA approach that suits their research aims, and automate the data analysis process. The framework includes four sequential steps: selecting a QCA approach, collecting and preprocessing mind maps, automated analysis, and validation, reliability, and model evaluation.

This framework aims to automate the generation of analytics with which to improve the effectiveness of EE. The automation framework makes analyzing mind maps used in EE research easier and scalable to large cohorts, more consistent and revealing, and capable of being used to evaluate differences among sample groups.

The automation framework has been tested in two experiments. The experiments' results have statistically supported the hypotheses of the process automation. First, the performance of classification models to automate deductive qualitative analysis of mind maps was similar to that of human analysis; and second, the performance of STM to automate the inductive qualitative analysis of mind maps was similar to that of human analysis. STM was also proven to outperform the standard topic modeling of LDA when both operated on the same data. STM generated latent topics that were statistically comparable to human categories, while LDA failed to produce the same quality of topics.

Exploiting topology of mind maps to improve the performance of process automation has been examined. In automating the deductive analysis, the use of map topology to extract features did not improve text classification performance. On the other hand, treating nodes as the unit of analysis when applying STM to automate inductive analysis has improved STM performance.

The two applications implemented as parts of the automation framework experimentation have demonstrated a successful attempt to apply state-of-the-art machine learning into EE research. The automation framework offers a unique and advanced qualitative research design that can be employed by EE researchers to benefit the EE best practices. The automation framework can enhance EE qualitative research in extracting textual statistical inference, shortening labor and time required by the analysis, measuring entrepreneurial motivations with machine learning and NLP techniques, increasing sample size, and ensuring validation and generalizability.

6.4 Suggestions for Future Research & Limitation

For future research, mind maps can include more than words. Students can insert pictures, use color, change the thickness of lines, add boxes, or graphics to make their maps unique and expressive (Buzan & Buzan, 2006). These potential elements require specific process automation. The accomplishment of such process automation could make using mind maps a much attractive data collection tool.

One of the automation framework advantages is the unconstrained increase in sample size. Although the sample size in this dissertation has matched those found in comparative qualitative studies, future research can involve a larger sample size. The sample size in this research has allowed machine learning techniques, classification models and STM, to achieve statistically significant results compared to human analysis, however, the automation framework can deal with massive samples.

The limitations of this research include the manual conversion of mind maps' content into plain text files. Using mind mapping software that enables analysts to export mind maps as plain text or XML files can solve this issue. For example, in R, a library '*xml*' allows analysts to read mind maps directly into the software.

The experiments to test the exploitation of mind map topology in the process automation were only performed on 12 datasets, including different feature representations of the motivation and deterrence datasets.

APPENDIX A

CLASSIFIERS A AND B COMPARISON

Table A.1 shows the F-score of classifiers "A" and "B," where classifier B was applied into the same feature representations but with adding mind map topology features under method two, i.e., specific-link analysis. It can be seen from the results that classifier B under the second method has improved classifier's B F-score in 9 datasets out of 12; however, this improvement overall was not statistically significant.

Table A.1 F-score of Classifiers A and B Under Method Two

DATASET	A	B
M-NPMI	0.4991	0.5632
M-TF	0.5756	0.6576
M-TFIDF	0.5224	0.5707
M-SVD-50	0.4283	0.4450
M-SVD-100	0.4744	0.4595
M-SVD-200	0.4904	0.5659
D-NPMI	0.6516	0.5416
D-TF	0.6966	0.5927
D-TFIDF	0.6149	0.6904
D-SVD-50	0.4769	0.4790
D-SVD-100	0.4966	0.5107
D-SVD-200	0.5432	0.6358

APPENDIX B

SRUCTURAL TOPIC MODEL

For nodes that were manually categorized as "Manager," Table B.1 shows that topic 15, 30, and 21 have been frequently assigned by STM as top topics, while topic 34 and 28 have interchangeably reserved the second and third top topics, while topic 27 and topic 19 have dominated fourth and fifth top topics, respectively. The table demonstrates how STM generates consistent topics to the same manual category.

Table B.1 Top Five Topics Inside A Sample of "Manager" Documents

ID	Category	1st Topic	2nd Topic	3rd Topic	4th Topic	5th Topic
FMa	Manager	Topic.15	Topic.34	Topic.28	Topic.27	Topic.19
FMb	Manager	Topic.30	Topic.34	Topic.28	Topic.27	Topic.19
FMb	Manager	Topic.21	Topic.28	Topic.34	Topic.27	Topic.19
FMc	Manager	Topic.21	Topic.34	Topic.28	Topic.27	Topic.19
FMe	Manager	Topic.15	Topic.34	Topic.28	Topic.27	Topic.19
FMf	Manager	Topic.15	Topic.28	Topic.34	Topic.27	Topic.14
FMk	Manager	Topic.15	Topic.34	Topic.28	Topic.27	Topic.19
FMy	Manager	Topic.30	Topic.34	Topic.28	Topic.27	Topic.19

Table B.2 shows a sample of documents with their top five topic-probability; each item represents the proportion of a specific topic inside a document. The difference in proportion between the top 1 and top 2 in one document is evident. It shows how STM has efficiently automated the inductive analysis of mind maps. Figures B.1 and B.2 show the correlation between topics and the topic's quality in the motivation documents.

Table B.2 Top Topic-Proportion of STM Inside A Sample of Documents

Document	Top 1	Top 2	Top 3	Top 4	Top 5
1	0.7203	0.0146	0.0140	0.0140	0.0095
2	0.8217	0.0090	0.0088	0.0088	0.0070
3	0.0455	0.0455	0.0451	0.0451	0.0436
4	0.5397	0.2623	0.0095	0.0095	0.0089
5	0.8207	0.0092	0.0088	0.0088	0.0069
6	0.6480	0.0171	0.0162	0.0162	0.0132
7	0.7020	0.0150	0.0145	0.0145	0.0115
8	0.8149	0.0095	0.0093	0.0093	0.0070
9	0.8677	0.0066	0.0065	0.0065	0.0050
10	0.5544	0.2483	0.0096	0.0096	0.0089
11	0.8139	0.0093	0.0092	0.0092	0.0070
12	0.2445	0.2445	0.2445	0.2445	0.0126
13	0.3846	0.3656	0.0123	0.0123	0.0098
14	0.1875	0.1875	0.1875	0.1875	0.0211
15	0.8207	0.0089	0.0085	0.0085	0.0069
16	0.8154	0.0092	0.0087	0.0087	0.0071
17	0.8160	0.0092	0.0091	0.0091	0.0070
18	0.7020	0.0150	0.0145	0.0145	0.0115
19	0.8525	0.0075	0.0070	0.0070	0.0053
20	0.8679	0.0066	0.0065	0.0065	0.0050
21	0.8217	0.0090	0.0088	0.0088	0.0070

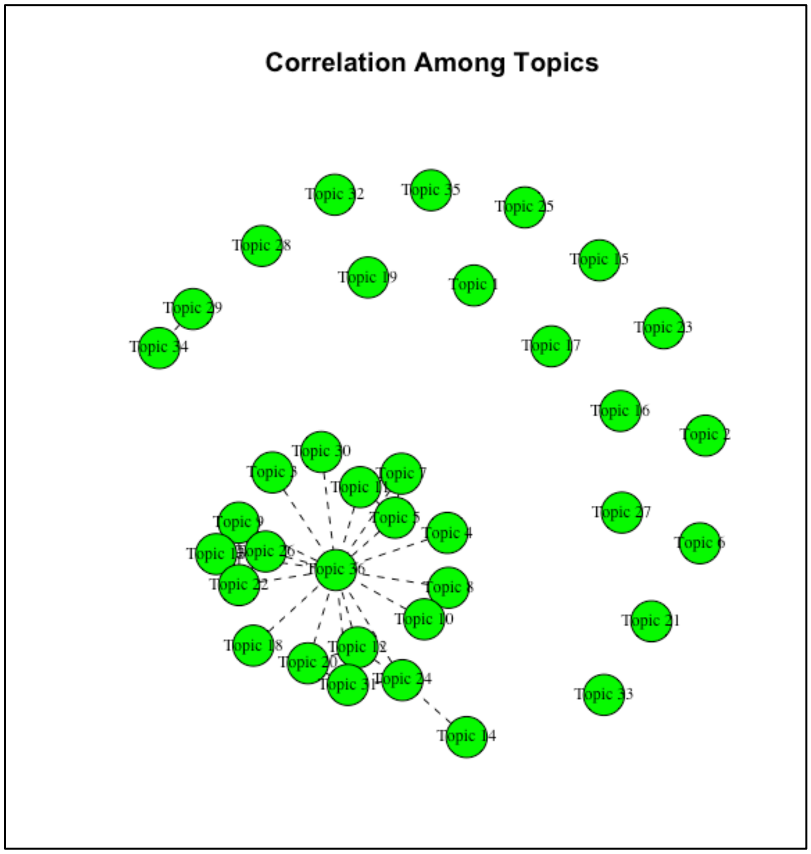

Figure B.1 Topics correlation – motivation.

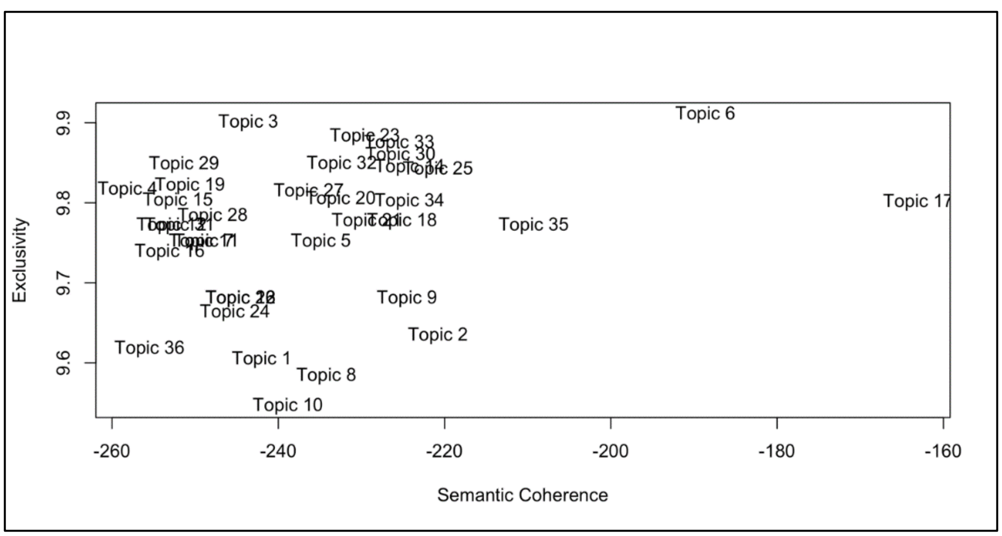

Figure B.2 Topics quality measured by semantic coherence and exclusivity for motivation mind maps with k = 36 STM.

REFERENCES

Acs, Z. J., & Audretsch, D. B. (2010). Introduction to the 2nd Edition of the Handbook of Entrepreneurship Research. In *Handbook of Entrepreneurship Research* (pp. 1-19), New York, NY: Springer.

Aggarwal, C. C., & Zhai, C. (2012). *Mining Text Data,* New York, NY: Springer Science & Business Media.

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician, 52*(2), 119-126.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Akinoglu, O., & Yasar, Z. (2007). The effects of note taking in science education through the mind mapping technique on students'attitudes, academic achievement and concept learning. *Journal of Baltic Science Education, 6*(3).

Aktinson, P., & Hammersley, M. (1998). Ethnography and participant observation. *Strategies of Qualitative Inquiry.* Thousand Oaks, CA: Sage, 248-261.

Alpaydin, E. (2009). *Introduction to Machine Learning*, Cambridge, Massachusetts : MIT Press.

Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (p. 485–499). Sage Publications, Inc.

Altheide, D. L., & Schneider, B. (1996). Process of qualitative document analysis. *Qualitative Media Analysis*, 23-41.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40-79.

Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., ... & Zhu, M. (2013, February). A practical algorithm for topic modeling with provable guarantees. In *International conference on machine learning,* Atlanta, Georgia (pp. 280-288).

Bae, T. J., Qian, S., Miao, C., & Fiet, J. O. (2014). The relationship between entrepreneurship education and entrepreneurial intentions: A meta–analytic review. *Entrepreneurship Theory and Practice, 38*(2), 217-254.

Bagozzi, B. E., & Berliner, D. (2018). The politics of scrutiny in human rights monitoring: Evidence from structural topic models of US State Department human rights reports. *Political Science Research and Methods, 6*(4), 661-677.

Bahr, H. M., Albrecht, S. L., & Chadwick, B. A. (1984). *Social Science Research Methods*, Upper Side River, New Jersey: Prentice Hall.

Bandera, C., Eminet, A., Passerini, K., & Pon, K. (2018a). 6. Cross-cultural entrepreneurship education: localization amidst globalization. *Annals of Entrepreneurship Education and Pedagogy–2018*, 99.

Bandera, C., Eminet, A., Passerini, K., & Pon, K. (2018b). Using mind maps to distinguish cultural norms between French and United States entrepreneurship students. *Small Business Management,* 56(S1), 177-196.

Barakat, S., Boddington, M., & Vyakarnam, S. (2014). Measuring entrepreneurial self-efficacy to understand the impact of creative activities for learning innovation. *The International Journal of Management Education, 12*(3), 456-468.

Basu, A., & Virick, M. (2008). Assessing entrepreneurial intentions amongst students: A comparative study. In *VentureWell. Proceedings of open, the annual conference* (p. 79): National Collegiate Inventors & Innovators Alliance.

Bauer, J., Qualmann, J., Stadtmüller, G., & Bauer, H. (1998). Lebenslauf untersuchungen bei alzheimer-patienten: Qualitative Inhaltsanalyse prämorbider Entwicklungsprozesse. *Psychosoziale Gerontologie, 2*, 251-274.

Baum, J. R., Olian, J. D., Erez, M., Schnell, E. R., Smith, K. G., Sims, H. P., . . . Smith, K. A. (1993). Nationality and work role interactions: A cultural contrast of Israeli and US entrepreneurs' versus managers' needs. *Journal of Business Venturing, 8*(6), 499-512.

Baumer, E. P., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology, 68*(6), 1397-1410.

Beck, K., & Vowe, G. (1995). Multimedia aus der Sicht der Medien. Argumentationsmuster und Sichtweisen in der medialen Konstruktion. *Rundfunk und Fernsehen, 43*(4), 549-563.

Beel, J., & Gipp, B. (2010, January). Link analysis in mind maps: a new approach to determining document relatedness. In *Proceedings of the 4th international conference on uniqiutous information management and communication*, Suwon Republic of Koria. (pp. 1-5)

Beel, J., Gipp, B., & Stiller, J.-O. (2009, November). Information retrieval on mind maps-what could it be good for*?* In *the 2009 5th International conference on collaborative computing*, Washington, DC: Networking: Applications and Worksharing.

Beel, J., & Langer, S. (2011, September). An exploratory analysis of mind maps. In *Proceedings of the 11th ACM symposium on document engineering*, Mountain View, CA. (pp. 81-84)

Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah? *Acm Sigkdd Explorations Newsletter, 2*(2), 1-13.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *J. Open Source Software, 3*(30), 774.

Berelson, B. (1959). The state of communication research. *Public Opinion Quarterly, 23*(1), 1-2.

Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage? In *the Proceedings of the 19th ACM international conference on information and knowledge management*, Tornto, Canada.

Bischof, J., & Airoldi, E. M. (2012, June). Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th international conference on machine learning (ICML-12)*, Madison, WI.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, New York, NY: Springer Science+ Business Media.

Blei, D. M. (2012). Surveying a suite of algorithms that offer a solution to managing large document archives. *Communication of the ACM, 55*(4), 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022.

Blenker, P., Trolle, E. S., Hedeboe, F. S., Korsgaard, S., & Wagner, K. (2014). Methods in entrepreneurship education research: a review and integrative framework. *Education+ Training, 56*(8/9), 697-715.

Bloehdorn, S., & Hotho, A. (2004, August). Boosting for text classification with semantic features. In *the International workshop on knowledge discovery on the web*, Seattle, WA.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1-2), 245-271.

Boissin, J.-P., Branchet, B., Emin, S., & Herbert, J. I. (2009). Students and entrepreneurship: a comparative study of France and the United States. *Journal of Small Business & Entrepreneurship, 22*(2), 101-122.

Bolton, D. L., & Lane, M. D. (2012). Individual entrepreneurial orientation: Development of a measurement instrument. *Education+ Training, 54*(2/3), 219-233.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the German society for computational linguistics and language technology*, Potsdam, Germany, 31-40.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145-1159.

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

Budd, J. W. (2004). Mind maps as classroom exercises. *The Journal of Economic Education, 35*(1), 35-46.

Burgess-Allen, J., & Owen-Smith, V. (2010). Using mind mapping techniques for rapid qualitative data analysis in public participation processes. *Health Expectations, 13*(4), 406-415.

Burla, L., Knierim, B., Barth, J., Liewald, K., Duetz, M., & Abel, T. (2008). From text to codings: intercoder reliability assessment in qualitative content analysis. *Nursing Research, 57*(2), 113-117.

Buzan, T., & Buzan, B. (2006). *The Mind Map Book*, London, UK: Pearson Education.

Cacciotti, G., Hayton, J. C., Mitchell, J. R., & Giazitzoglu, A. (2016). A reconceptualization of fear of failure in entrepreneurship. *Journal of Business Venturing, 31*(3), 302-325.

Calle, M. L., & Urrea, V. (2011). Letter to the editor: stability of random forest importance measures. *Briefings in Bioinformatics, 12*(1), 86-89.

Carley, K., & Roberts, C. W. (1997). *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences From Texts and Transcripts*, Mahwah, NJ: Erlbaum.

Carsrud, A., & Brännback, M. (2011). Entrepreneurial motivations: What do we still need to know? *Journal of Small Business Management, 49*(1), 9-26.

Chang, J., & Chang, M. J. (2010). Package ‘lda’. *In: Citeseer*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.216.2273&rep=rep1&type=pdf

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009, December). Reading tea leaves: How humans interpret topic models. In *the 23rd Advances in neural information processing systems*, Vancouver, Canada.

Chen, X.-P., Yao, X., & Kotha, S. (2009). Entrepreneur passion and preparedness in business plan presentations: a persuasion analysis of venture capitalists' funding decisions. *Academy of Management Journal, 52*(1), 199-214.

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). Btm: topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering, 26*(12), 2928-2941.

Jurka, T. P., Collingwood, L., Boydstun, A. E., Grossman, E., & van Atteveldt, W. (2013). RTextTools: A Supervised Learning Package for Text Classification. *R Journal*, 5(1).

Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory,* Thousand Oaks, CA: Sage publications.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Crook, T. R., Shook, C. L., Morris, M. L., & Madden, T. M. (2010). Are we there yet? An assessment of research design and construct measurement practices in entrepreneurship research. *Organizational Research Methods, 13*(1), 192-206.

Curth, A., Chatzichristou, S., Devaux, A., & Allinson, R. (2015). Entrepreneurship education: a road to success. A Compilation of evidence on the impact of entrepreneurship education strategies and measures. Belgium: European Commission.

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher Education, 62*(3), 279-301.

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *the Proceedings of the 23rd international conference on machine learning*, Pittsburgh, Pennsylvania.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*(Jan), 1-30.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 20*(1), 134-144.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87.

Dumais, S., & Chen, H. (2000, July). Hierarchical classification of web content. In *the Proceedings of the 23rd annual international ACM special interest group on information retrieval conference on research and development in information retrieval,* Athens, Greece.

Duval-Couetil, N., Reed-Rhoads, T., & Haghighi, S. (2012). Engineering students and entrepreneurship education: Involvement, attitudes and outcomes. *International Journal of Engineering Education, 28*(2), 425.

Duval-Couetil, N. (2013). Assessing the impact of entrepreneurship education programs: Challenges and approaches. *Journal of Small Business Management, 51*(3), 394-409.

Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., & Stewart, B. M. (2018). How to make causal inferences using texts. *arXiv preprint arXiv:1802.02163*.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107-115.

Eppler, M. J. (2006). A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization, 5*(3), 202-210.

Erichson, N. B., Voronin, S., Brunton, S. L., & Kutz, J. N. (2016). Randomized matrix decompositions using R. *arXiv preprint arXiv:1608.02148*.

Esses, V. M., & Maio, G. R. (2002). Expanding the assessment of attitude components and structure: The benefits of open-ended measures. *European Review of Social Psychology, 12*(1), 71-101.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research, 9*(Aug), 1871-1874.

Farhangmehr, M., Gonçalves, P., & Sarmento, M. (2016). Predicting entrepreneurial motivation among university students: The role of entrepreneurship education. *Education + Training, 58*(7/8). doi:10.1108/ET-01-2016-0019

Fayolle, A. (2013). Personal views on the future of entrepreneurship education. *Entrepreneurship & Regional Development, 25*(7-8), 692-701.

Fayolle, A., Gailly, B., & Lassas-Clerc, N. (2006). Assessing the impact of entrepreneurship education programmes: a new methodology. *Journal of European industrial training, 30*(9), 701-720.

Feinerer, I. (2013). Introduction to the tm Package Text Mining in R. *Accessible en ligne: http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf. Retrieved March, 1*, 2019.

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data,* Cambridge, UK: Cambridge University Press.

Fiet, J. O. (2001). The theoretical side of teaching entrepreneurship. *Journal of Business Venturing, 16*(1), 1-24.

Franco, M., Haase, H., & Lautenschläger, A. (2010). Students' entrepreneurial intentions: an inter-regional comparison. *Education+ Training, 52*(4), 260-275.

Freedman, D. A. (2009). *Statistical Models: Theory and Practice,* Cambridge, UK: Cambridge University Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning*, New York, NY: Springer series in statistics New York.

Gadarian, S. K., & Albertson, B. (2014). Anxiety, immigration, and the search for information. *Political Psychology, 35*(2), 133-164.

Gartner, W. B., & Birley, S. (2002). Introduction to the special issue on qualitative methods in entrepreneurship research. In: Elsevier.

Geer, J. G. (1988). What do open-ended questions measure? *Public Opinion Quarterly, 52*(3), 365-367.

Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics, 49*(3), 291-304.

Giacomin, O., Janssen, F., Pruett, M., Shinnar, R. S., Llopis, F., & Toney, B. (2011). Entrepreneurial intentions, motivations and barriers: Differences among American, Asian and European students. *International Entrepreneurship and Management Journal, 7*(2), 219-238.

Giorgi, A. (2009). *The Descriptive Phenomenological Method in Psychology: A Modified Husserlian Approach*, Pittsburgh, PA: Duquesne University Press.

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies, 10*(1), 1-309.

Golub, G. H., & Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Linear Algebra* (pp. 134-151): Springer.

Grajzl, P., & Murrell, P. (2019). Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas. *Journal of Comparative Economics, 47*(1), 111-135.

Graneheim, U. H., Lindgren, B.-M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today, 56*, 29-34.

Greenwood, P. E., & Nikulin, M. S. (1996). *A Guide to Chi-squared Testing,* (Vol. 280), Hoboken, NJ: John Wiley & Sons.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promice and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267-297.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature Extraction: Foundations and Applications,* New York, NY, (Vol. 207): Springer.

Haddock, G., & Zanna, M. P. (1998). On the use of open-ended measures to assess attitudinal components. *British Journal of Social Psychology, 37*(2), 129-149.

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Management, 54*(6), 1292-1307.

Han, S., Qubo, C., & Meng, H. (2012, June). Parameter selection in SVM with RBF kernel function. In *World Automation Congress* 2012 (pp. 1-4). IEEE.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing, 36*(1), 20-38.

Hayton, J. C., George, G., & Zahra, S. A. (2002). National culture and entrepreneurship: A review of behavioral research. *Entrepreneurship Theory and Practice, 26*(4), 33-52.

Hickey, G., & Kipping, C. (1996). A multi-stage approach to the coding of data from open-ended questions. *Nurse Researcher, 4*(1), 81-91.

Hillard, D., Purpura, S., & Wilkerson, J. (2007). An active learning framework for classifying political text. In *the Annual Meeting of the Midwest Political Science* Association, Chicago.

Hinton, G. E., Sejnowski, T. J., & Poggio, T. A. (1999). *Unsupervised Learning: Foundations of Neural Computation,* Cambridge, Massachusetts: MIT Press.

Hofstede, G. (1983). Cultural dimensions for project management. *International Journal of Project Management, 1*(1), 41-48.

Hofstede, G. (1994). The business of international business is culture. *International Business Review, 3*(1), 1-14.

Hofstede, G. (2003). Cultural dimensions. *www. geert-hofstede. com*. "accessed on December 2nd, 2019"

Hofstede, G., Noorderhaven, N. G., Thurik, A. R., Uhlaner, L. M., Wennekers, A. R., & Wildeman, R. E. (2004). Culture's role in entrepreneurship: self-employment out of dissatisfaction. *Innovation, Entrepreneurship and Culture: The Interaction between Technology, Progress and Economic Growth, 162203*.

Hoftede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and its Importance for Survival,* New York, NY: McGraw-Hill.

Hope, A. C. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society: Series B (Methodological), 30*(3), 582-598.

Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software, 40*(13), 1-30.

Horton, P. B., McConney, A. A., Woods, A. L., Barry, K., Krout, H. L., & Doyle, B. K. (1993). A content analysis of research published in the journal of research in science teaching from 1985 through 1989. *Journal of Research in Science Teaching, 30*(8), 857-869.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398), Hoboken, NJ: John Wiley & Sons.

Hsieh, H. F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277-1288.

Hsu, D. K., Shinnar, R. S., & Powell, B. C. (2014). Expectancy theory and entrepreneurial motivation: A longitudinal examination of the role of entrepreneurship education. *Journal of Business and Entrepreneurship, 26*(1), 121-140.

Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus, 5*(1), 1-9.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference* (NZCSRSC2008), Christchurch, New Zealand.

Humphreys, A., & Wang, R. J.-H. (2017). Automated text analysis for consumer research. *Journal of Consumer Research, 44*(6), 1274-1306.

Hurschka, D. J., Schwartz, D., John, D. C. S., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods, 15*(9), 1277-1288.

Hynes, B. (1996). Entrepreneurship education and training-introducing entrepreneurship into non-business disciplines. *Journal of European Industrial Training, 20*(8), 10-17.

Iyengar, S. (1996). Framing responsibility for political issues. *The Annals of the American Academy of Political and Social Science, 546*(1), 59-70.

Jackson, K. M., & Trochim, W. M. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods, 5*(4), 307-336.

Jamieson, P. (2012, October). Using modern graph analysis techniques on mind maps to help quantify learning. In *the 2012 Frontiers in Education Conference Proceedings,* Seattle, WA.

Jivani, A. G. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl, 2*(6), 1930-1938.

Joachims, T. (1998). Making large-scale SVM learning practical (No. 1998, 28). *Technical Report*.

Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education, 13*(2), 121-138.

Jones, B. D., Ruff, C., Snyder, J. D., Petrich, B., & Koonce, C. (2012). The effects of mind mapping activities on students' motivation. *International Journal for the Scholarship of Teaching and Learning, 6*(1).

Katz, J. A. (2003). The chronology and intellectual trajectory of American entrepreneurship education: 1876–1999. *Journal of Business Venturing, 18*(2), 283-300.

Kern, C. S., Bush, K. L., & McCleish, J. M. (2006). Mind-mapped care plans: Integrating an innovative educational tool as an alternative to traditional care plans. *Journal of Nursing Education, 45*(4).

Kevany, K., Huisingh, D., García, F. J. L., Lourdel, N., Gondran, N., Laforest, V., . . . Brodhag, C. (2007). Sustainable development cognitive map: a new method of evaluating student understanding. *International Journal of Sustainability in Higher Education*.

Kim-Soon, N., Ahmad, A. R., & Ibrahim, N. N. (2016). Theory of planned behavior: undergraduates' entrepreneurial motivation and entrepreneurship career intention at a public university. *Journal of Entrepreneurship: Research & Practice, 2016*, 1-14.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic Regression,* New York, NY: Springer.

Kolvereid, L. (1996). Prediction of employment status choice intentions. *Entrepreneurship Theory and Practice, 21*(1), 47-58.

Kolvereid, L., & Moen, Ø. (1997). Entrepreneurship among business graduates: does a major in entrepreneurship make a difference? *Journal of European Industrial Training, 21*(4), 154-160.

Kondracki, N. L., Wellman, N. S., & Amundson, D. R. (2002). Content analysis: Review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behavior, 34*(4), 224-230.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering, 160*, 3-24.

Kracauer, S. (1952). The challenge of qualitative content analysis. *Public Opinion Quarterly*, 631-642.

Krippendorff, K. (2018). *Content analysis: An Introduction to Its Methodology,* Thousand Oaks, CA: Sage publications.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411-433.

Krueger, R. A. (2006). Analyzing focus group interviews. *Journal of Wound Ostomy & Continence Nursing, 33*(5), 478-481.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models,* Irwin, NY, (Vol. 5): McGraw-Hill.

Labatut, V., & Cherifi, H. (2012). Accuracy measures for the comparison of classifiers. *arXiv preprint arXiv:1207.3790*.

Lau, J. H., Newman, D., Karimi, S., & Baldwin, T. (2010, August). Best topic word selection for topic labelling. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China: Posters.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.

Lee, S. I., Lee, H., Abbeel, P., & Ng, A. Y. (2006, July). Efficient l~ 1 regularized logistic regression. In *Association for the Advancement of Artificial Intelligence* (Vol. 6, pp. 401-408).

Lee, S. M., & Peterson, S. J. (2000). Culture, entrepreneurial orientation, and global competitiveness. *Journal of World Business, 35*(4), 401-416.

Levenburg, N. M., Lane, P. M., & Schwarz, T. V. (2006). Interdisciplinary dimensions in entrepreneurship. *Journal of Education for Business, 81*(5), 275-281.

Lewis, D. D., & Gale, W. A. (1994). *A sequential algorithm for training text classifiers.* Paper presented at the SIGIR'94.

Lichy, J., & Pon, K. (2015). For better or for worse: the changing life of academic staff in French business schools. *Journal of Management Development, 34*(5), 536-552.

Liguori, E., Cowden, B., & Hertz, G. (2016). Teaching entrepreneurial sales skills: a co-curricular approach. In *Annals of Entrepreneurship Education and Pedagogy–2016*. Edward Elgar Publishing.

Liguori, E., Winkler, C., Winkel, D., Marvel, M. R., Keels, J. K., van Gelderen, M., & Noyes, E. (2018). The entrepreneurship education imperative: Introducing EE&P. In *SAGE Publications*, Los Angeles, CA.

Lincoln, Y. S., & Guba, E. G.(1985). *Naturalistic Inquiry*, Beverly Hills, CA: Sage.

Lindkvist, K. (1981). Approaches to textual analysis. *Advances in Content Analysis, 9*, 23-42.

Liu, X., Lin, C., Zhao, G., & Zhao, D. (2019). Research on the Effects of Entrepreneurial Education and Entrepreneurial Self-Efficacy on College Students' Entrepreneurial Intention. *Frontiers in Psychology, 10*.

Lorz, M., Mueller, S., & Volery, T. (2013). Entrepreneurship education: a systematic review of the methods in impact studies. *Journal of Enterprising Culture, 21*(02), 123-151.

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis, 23*(2), 254-277.

Mankad, S., Han, H. S., Goh, J., & Gavirneni, S. (2016). Understanding online hotel reviews through automated text analysis. *Service Science, 8*(2), 124-138.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing,* Campridge, MA: MIT press.

Maresch, D., Harms, R., Kailer, N., & Wimmer-Wurm, B. (2016). The impact of entrepreneurship education on the entrepreneurial intention of students in science and engineering versus business studies university programs. *Technological Forecasting and Social Change, 104*, 172-179.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems, 54*(1), 11-22.

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice, 13*(6), 522-526.

Martin, B. C., McNally, J. J., & Kay, M. J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing, 28*(2), 211-224.

Matlay, H., Taatila, V., & Down, S. (2012). Measuring entrepreneurial orientation of university students. *Education+ Training*.

Mayring, P. (2004). Qualitative content analysis. *A Companion to Qualitative Research, 1*, 159-176.

Mayring, P. (2014). Qualitative content analysis: theoretical foundation, basic procedures and software solution. "accessed on January, 15th, 2020"

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica: Biochemia Medica, 23*(2), 143-149.

McTavish, D. G., & Pirro, E. B. (1990). Contextual content analysis. *Quality and Quantity, 24*(3), 245-265.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013, December). Distributed representations of words and phrases and their compositionality. In *the Advances in neural information processing systems,* Harrah's Lake Tahoe, NA.

Mimno, D., & Lee, M. (2014, October). Low-dimensional embeddings for interpretable anchor-based topic inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP), Doha, Qatar.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011, July). *Optimizing semantic coherence in topic models.* In *Proceedings of the conference on empirical methods in natural language processing*, Edinburgh, Scotland, UK.

Molina-Azorín, J. F., López-Gamero, M. D., Pereira-Moliner, J., & Pertusa-Ortega, E. M. (2012). Mixed methods studies in entrepreneurship research: Applications and contributions. *Entrepreneurship & Regional Development, 24*(5-6), 425-456.

Moore, C. F. (1986, December). Understanding entrepreneurial behavior: A definition and model. In the Academy of Management Proceedings, (pp. 66-70).

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications, 40*(2), 621-633.

Moroz, P. W., & Hindle, K. (2012). Entrepreneurship as a process: Toward harmonizing multiple perspectives. *Entrepreneurship Theory and Practice, 36*(4), 781-818.

Morris, M. H., & Liguori, E. (2016). Preface: Teaching reason and the unreasonable. *Annals of Entrepreneurship Education and Pedagogy, 2*, xiv-xxii.

Mostyn, B. (1985). The content analysis of qualitative research data, a dynamic approach. *The Research Interview, Users and Approaches*, 115-145.

Mueller, S. L., & Thomas, A. S. (2001). Culture and entrepreneurial potential: A nine country study of locus of control and innovativeness. *Journal of Business Venturing, 16*(1), 51-75.

Mullen, M. R., Budeva, D. G., & Doney, P. M. (2009). Research methods in the leading small business–entrepreneurship journals: A critical review with recommendations for future research. *Journal of Small Business Management, 47*(3), 287-307.

Murnieks, C. Y., Klotz, A. C., & Shepherd, D. A. (2020). Entrepreneurial motivation: A review of the literature and an agenda for future research. *Journal of Organizational Behavior, 41*(2), 115-143.

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology, 4*(1), 13-20.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1.

Neck, H. M., Greene, P. G., & Brush, C. G. (2014). *Teaching Entrepreneurship: A Practice-based Approach,* Cheltenham, UK: Edward Elgar Publishing.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics,* Los Angeles, CA.

Nuttin, J., Lorion, R. P., & Dumas, J. E. (1984). *Motivation, Planning, and Action: A Relational Theory of Behavior Dynamics,* Leuvem, Beljium: Leuven University Press.

Oosterbeek, H., Van Praag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review, 54*(3), 442-454.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1-135.

Pang, B. L., L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1-2), 1-135.

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice,* Thousand Oaks, CA: Sage publications.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14.

Pihie, Z. A. L., & Bagheri, A. (2011). Teachers' and students' entrepreneurial self-efficacy: Implication for effective teaching practices. *Procedia-Social and Behavioral Sciences, 29*, 1071-1080.

Pittaway, L., & Cope, J. (2007). Entrepreneurship education: A systematic review of the evidence. *International Small Business Journal, 25*(5), 479-510.

Popping, R. (2015). Analyzing open-ended questions by means of text analysis procedures. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 128*(1), 23-39. doi:10.1177/0759106315597389

Potter, W. J., & Levine-Donnerstein, D. (1999). *Rethinking Validity and Reliability in Content Analysis*, Milton Park, UK: Taylor & Francis.

Prasad, B. D. (2008). Content analysis. *Research Methods for Social Work, 5*, 1-20.

Preoţiuc-Pietro, D., Lampos, V., & Aletras, N. (2015). An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China (Volume 1: Long Papers).

Quan, X., Kit, C., Ge, Y., & Pan, S. J. (2015, July). Short and sparse text topic modeling via self-aggregation. In *24th International Joint Conference on Artificial Intelligence*. (pp. 2270-2276)

Rahman, M. S. (2017). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language" testing and assessment" research: A literature review. *Journal of Education and Learning, 6*(1), 102-112.

Ramya, R., Venugopal, K., Iyengar, S., & Patnaik, L. (2017). Feature extraction and duplicate detection for text mining: A survey. *Global Journal of Computer Science and Technology*.

Read, T. R., & Cressie, N. A. (2012). *Goodness-of-fit Statistics for Discrete Multivariate Data,* New York, NY: Springer Science & Business Media.

Reich, J., Tingley, D. H., Leder-Luis, J., Roberts, M., & Stewart, B. (2014). Computer-assisted reading and discovery for student-generated text in massive open online courses. *Journal of Learning Analytics*, 2(1), 156–184

RePass, D. E. (1971). Issue salience and party choice. *The American Political Science Review, 65*(2), 389-400.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). stm: R package for structural topic models. *Journal of Statistical Software, 10*(2), 1-40.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). Navigating the local modes of big data. *Computational Social Science, 51*.

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013). The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation* (Vol. 4).

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064-1082. doi:10.1111/ajps.12103

Ruiz-Primo, M. A., Schultz, S. E., & Shavelson, R. J. (1997). Concept map-based assessment in science: Two exploratory studies. *Center for Research on Evaluation, Standards, and Student Testing*, Graduate School of Education & Information Studies, University of California, Los Angeles.

Ruiz-Primo, M. A., Schultz, S. E., & Shavelson, R. J. (1997b). *On the Validity of Concept Map-base Assessment Interpretations: An Experiment Eesting the Assumption of Hierarchical Concept Maps in Science*, Los Angeles, CA: CRESST.

Salzberg, S. L. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery, 1*(3), 317-328.

Schank, R. C., & Abelson, R. P. (2013). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures,* Hove, UK: Psychology Press.

Scharkow, M. (2013). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity, 47*(2), 761-773.

Scheffe, H. (1999). *The Analysis of Variance,* (Vol. 72), Hoboken, NJ: John Wiley & Sons.

Schölkopf, B., Simard, P., Smola, A. J., & Vapnik, V. (1998). Prior knowledge in support vector kernels. In *Advances in neural information processing systems*. (pp. 640-646)

Scholkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond,* Cambridge, MA: MIT press.

Schreier, M. (2012). *Qualitative Content Analysis in Practice,* Thousand Oaks, CA: Sage publications.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR), 34*(1), 1-47.

Serry, T., & Liamputtong, P. (2013). The in-depth interviewing method in health. *Research Methods in Health: Foundations for Evidence-based Practice*, 39-53.

Shane, S., Locke, E. A., & Collins, C. J. (2003). Entrepreneurial motivation. *Human Resource Management Review, 13*(2), 257-279.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Shinnar, R., Pruett, M., & Toney, B. (2009). Entrepreneurship education: Attitudes across campus. *Journal of Education for Business, 84*(3), 151-159.

Shinnar, R. S., Giacomin, O., & Janssen, F. (2012). Entrepreneurial perceptions and intentions: The role of gender and culture. *Entrepreneurship Theory and Practice, 36*(3), 465-493.

Shor, E., Rijt, A. v. d., Miltsov, A., Kulkarni, V., & Skiena, S. (2015). A Paper ceiling: Explaining the persistent underrepresentation of women in printed news. *American Sociological Review, 80*(5), 960-984.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *J. Open Source Software, 1*(3), 37.

Singer, S., Amorós, J. E., & Moska, D. (2015). Global Entrepreneurship Monitor (GEM) 2014 global report. *Global Entrepreneurship Research Association,* London Business School, London.

Slonim, N., & Tishby, N. (2000, July). Document clustering using word clusters via the information bottleneck method. In *Proceedings of the 23rd annual international ACM conference on research and development in information retrieval,* Athens, Greece.

Soares, A. M., Farhangmehr, M., & Shoham, A. (2007). Hofstede's dimensions of culture in international marketing studies. *Journal of Business Research, 60*(3), 277-284.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437.

Solesvik, M. Z. (2013). Entrepreneurial motivations and intentions: investigating the role of education major. *Education+ Training*.

Solomon, G., & Matlay, H. (2008). The impact of entrepreneurship education on entrepreneurial outcomes. *Journal of Small Business and Enterprise Development*.

Somers, M. J., Passerini, K., Parhankangas, A., & Casal, J. (2014). Using mind maps to study how business school students and faculty organize and apply general business knowledge. *The International Journal of Management Education, 12*(1), 1-13.

Sosinsky, B. (2009). *Networking bible,* (Vol. 567), Hoboken, NJ: John Wiley & Sons.

Souitaris, V., Zerbinati, S., & Al-Laham, A. (2007). Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *Journal of Business Venturing, 22*(4), 566-591.

Stone, P. J., & Hunt, E. B. (1963, May). *A computer approach to content analysis: studies using the general inquirer system.* In Proceedings of the May 21-23, 1963, spring joint computer conference, New York, NY.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted Studies of Language and Culture,* Oxford, UK: Blackwell Oxford.

Tackey, N., Perryman, S., & Connor, H. (1999). Graduate business start-ups. *Institute for Employment Studies, UK*.

Taddy, M. (2012). On estimation and selection for topic models. In *Artificial Intelligence and Statistics*. (pp. 1184-1193)

Tattersall, C., Watts, A., & Vernon, S. (2007). Mind Mapping as a Tool in Qualitative Research. *Nursing Times, 103*(26), 32-33.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: Liwc and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24-54.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2005). Sharing clusters among related groups: Hierarchical Dirichlet processes. In *Advances in neural information processing systems*.

Tesch, R. (2013). *Qualitative Research: Analysis Types and Software,* Abingdon, UK: Routledge.

Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research, 15*(1), 3221-3245.

Vanevenhoven, J., & Liguori, E. (2013). The Impact of Entrepreneurship Education: Introducing the E ntrepreneurship E ducation P roject. *Journal of Small Business Management, 51*(3), 315-328.

Verheul, I., Wennekers, S., Audretsch, D., & Thurik, R. (2002). An eclectic theory of entrepreneurship: policies, institutions and culture. In *Entrepreneurship: determinants and policy in a European-US comparison* (pp. 11-81): Springer.

Vesper, K. H., & Gartner, W. B. (1997). Measuring progress in entrepreneurship education. *Journal of Business Venturing, 12*(5), 403-421.

Vicini, S. (1993). *Subjektive Beratungstheorien: bernische ErziehungsberaterInnen reflektieren ihre Praxis*: Lang.

Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In A practical approach to microarray data analysis (pp. 91-109): Springer.

Wallach, H. M. (2006, June). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on machine learning*, New York, NY. (pp. 977-984)

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). *Evaluation methods for topic models.* In Proceedings of the 26th annual international conference on machine learning, Montreal Quebec, Canada.

Weber, R. P. (1990). *Basic Content Analysis,* Thousand Oaks, CA: Sage.

West, D. C., Park, J. K., Pomeroy, J. R., & Sandoval, J. (2002). Concept mapping assessment in medical education: a comparison of two scoring systems. *Medical Education, 36*(9), 820-826.

Wheeldon, J. (2010). Mapping mixed methods research: Methods, measures, and meaning. *Journal of Mixed Methods Research, 4*(2), 87-102.

Wheeldon, J., & Ahlberg, M. (2012). Mapping mixed-methods research: Theories, models, and measures. *Visualizing Social Science Research: Maps, Methods, & Meaning*, 113-148.

Wheeldon, J., & Ahlberg, M. (2017). Mind maps in qualitative research. *Handbook of Research Methods in Health Social Sciences*, 1-17.

Wheeldon, J., & Faubert, J. (2009). Framing experience: Concept maps, mind maps, and data collection in qualitative research. *International Journal of Qualitative Methods, 8*(3), 68-83.

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends, 55*(1), 22-45.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in Statistics* (pp. 196-202): Springer.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 2*(1-3), 37-52.

Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* (p. 217–244). American Psychological Association.

Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS), 26*(3), 13.

Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval, 1*(1-2), 69-90.

Yang, Y., & Liu, X. (1999, August). A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM conference on research and development in information retrieval,* Berkeley, CA.

Yu, W., Ruibo, W., Huichen, J., & Jihong, L. (2014). Blocked 3× 2 cross-validated t-test for comparing supervised classification learning algorithms. *Neural Computation, 26*(1), 208-235.

Zahra, S. A. (2007). Contextualizing theory building in entrepreneurship research. *Journal of Business Venturing, 22*(3), 443-452.

Zampetakis, L. A., Tsironis, L., & Moustakis, V. (2007). Creativity development in engineering education: The case of mind mapping. *Journal of Management Development, 26*(4), 370-380.

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics, 1*(1-4), 43-52.

Zuo, Y., Zhao, J., & Xu, K. (2016). Word network topic model: a simple but general solution for short and imbalanced texts. *Knowledge and Information Systems, 48*(2), 379-398.