Text-mining and Analysis of the Doctors' Meta-data and Text-reviews using Topic-modeling (LDA) Technique

by

Asad Khan

Submitted in partial fulfilment of the requirements
for the degree of Master of Electronic Commerce

at

Dalhousie University
Halifax, Nova Scotia
September 2019

DEDICATION PAGE

I am dedicating this thesis to six cherished people who have meant and continue to mean so much to me. First and foremost, to my father, who is the source of sheer inspiration in my life and who taught me the value of hard work. To my mother - mom your sweet memories will always be guiding my life. I will never forget you.

Next, my beloved wife, Mrs Nazia Asad, who stood by me through the ups and downs of life and particularly the period of conducting this research.

I also want to thank my dearest children, Anousha Asad, Rida Asad and Yahya Saud Khan for sparing me not been able to spend time with them during my studies.

I love you all beyond words.

Asad Khan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The developments in the internet and web technologies along with smart devices have empowered consumers to rate, comment, review, recommend products and services for others using a plethora of platforms, such as RateMDs.com. Therefore, feedback is critical to improve the overall quality of a process, product or service. Hence, the healthcare industry is no exception. This thesis aims to mine and analyze physicians' online reviews using web-scrapping and topic-modeling (LDA) technique. RateMDs.com was chosen as a case study for the period from September 2013 to January 2019. The thesis employed web scrapping, to collect physicians' meta-data, and LDA technique, a generative probabilistic model of text-corpus to the text-corpus, for text-mining among Canadian provinces. The results revealed that physicians, in some of the specialities, such as plastic surgery, had a higher probability of being rated than others in specialities such as Radiation, Oncology and Osteopathy. The research also revealed that East coast provinces had a relatively higher rating than those in the West of Canada. Finally, this thesis validates the use of Python (BeautifulSoup, spaCy, Gensim, NLTK, re) for text-mining with LDA.

Keywords: Physician Rating, KPI (Key Performance Indicator), Machine Learning, Unsupervised Learning, Text-mininng, Topic Modeling, Latent Dirichlet Allocation (LDA)

LIST OF ABBREVIATIONS USED

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DNLP	Dalhousie Natural Language Processing
HAL	Hierarchical Anticipatory Learning
HCI	Human Computer Interaction
IPYNB	Interactive Jupyter Python Notebook
LDA	Latent Dirichlet Method
ML	Machine Learning
NLP	Natural Language Processing
PY	Python

ACKNOWLEDGEMENTS

First and foremost, I want to acknowledge my thesis supervisor Dr. Rita Orji, Faculty of Computer Science at Dalhousie University. She has always been helpful and instrumental in the completion of my thesis. She consistently guided me to ensure my success by steering me in the right direction.

Besides my supervisr, I would like to thank the rest of my thesis committee: Dr Vlado Keselj, Dr Colin Conrad, Gabriella Mosquera, Dr Israat Haque (Chair) of the Faculty of Computer Science for their encouragement, insightful comments, and hard questions. I would also wish to acknowledge the involved experts of the Faculty of Computer Science who have supported me in the completion of this thesis: Dijana Kosmajac of Dalhousie Natural Language Processing Group (DNLP), Zaher Abdulmaula of Hierarchical Anticipatory Learning (HAL) Lab, Oladapo Oyebode of Human Computer Interaction (HCI) Lab of the Faculty of Computer Science, here at Dalhousie University. Their passionate participation was quite pivotal in the making of this thesis a success. I am thankful for the valuable suggestions of all the experts involved in this thesis.

Finally, I want to thank my parents and beloved wife – Nazia Asad for supporting and encouraging me each step of the way throughout my study years and mostly through my time of undertaking this research study. I would not have succeeded in accomplishing this project without their help. Thank you.

Asad Khan

CHAPTER 1 INTRODUCTION

1.1 MOTIVATION AND SCOPE

Feedback is one of the vital KPIs (Key Performance Indicators) for many business or service providers. Similarly, healthcare-providers rely on such feedback provided by their consumers using different online platforms to improve the level of their services. Websites that allow healthcare consumers to review their respective healthcare providers, such as healthcare providers, are well known today. Such physician rating and review websites have been around for more than ten years; however, the healthcare consumers feedback has often been underappreciated [1]. RateMDs.com is one of such websites that allow customers to provide online reviews of their healthcare providers.

This thesis uses the meta-data of doctors in Canada from RateMDs.com for the period September 2013 to January 2019. The meta-data is used to study the number of ratings, unique rated healthcare providers, rating per physician and overall rating of all physician rated by practice along with their proportion of mean by practice. In addition, text-reviews have been analyzed to detect the most recurring aspects using topic-modeling technique LDA. Python was used for web-scrapping and LDA model because Python is the de-facto standard for open-source data science projects and is widely used by academics and businesses.

1.2 PROBLEM STATEMENT

There are numerous research studies on the formal evaluation of healthcare providers in Canada, such as Liu et. al [9]. Learning more about the relationship between healthcare providers and consumers, using fully automated methods of mining text-reviews could help the healthcare authorities in improving the overall service quality. Furthermore, in order to be able to understand the complexities of doctors' online text-reviews, it is important that a case study be conducted to examine the most recurring issues based on doctors' specialties and geographical location.

1.3 RESEARCH AIMS

This thesis aims to investigate and analyze both the meta-data and text-reviews scraped from RateMDs.com for the period from 2013-2019. The following questions are set as guidelines to fulfil the aim:

(1). To what extent are the distinguishing salient features reflected in the reviews for the period September 2013 to January 2019 by specialties and by geographical region?

(2). What are the topics that are commonly discussed by raters in the healthcare providers' reviews?

1.4 RESEARCH OBJECTIVES

From the research questions above, this research had the following objectives:

(1) To determine the extent of changes in the online doctors' reviews for the period from September 2013 to January 2019.

(2) To apply the LDA (Latent Dirichlet Allocation) technique to detect the salient aspects from the text-reviews of certain specialties from RateMDs.com.

1.5 SOLUTION

To achieve the objective (1) I developed a web scrapping bot, using Python's BeautifulSoup, to collect the meta-data about the doctors/healthcare providers. Next, I analyzed the meta-data using descriptive statistics to uncover some insights. While, for the objective (2) I employed the Latent Dirichlet Allocation (LDA) topic-modeling techniques, a generative probabilistic model of text-corpus to the text-corpus. The results revealed that healthcare providers, in some of the specialities, such as plastic surgery, had a higher probability of being rated than others in specialities such as radiation oncology and osteopathy. In addition, this research study reveals that although there is insignificant differences in rating by geographical regions, Atlantic provinces had a relatively higher rating than those in the West of Canada. The LDA uncovered several interesting insights. The top two most negative tokens from the model analysis turned out to be "arrogant" and "annoying". They highlighted the importance of communication and interpersonal skills for healthcare providers and consumers relationships. Positive remarks, on the other hand, were dominated by superlatives such as "respectful" and "competent". Finally, this thesis validates the use of Python (BeautifulSoup, spaCy, Gensim, NLTK, re) for text-mining with LDA.

There are several practical benefits of comparing Provinces and Professions (specialties). The Healthcare in Canada is a provincial matter, which means that the healthcare management differs substantially. For instance, in Quebec Doctors serve patients on a case basis, i.e. one healthcare problem per visit. This research study explores insights related to the healthcare practitioners that could be used in collaboration with information from CIHI (Canadian Institute for Health Information) [3].

1.6 ETHICS APPROVAL

The research study is exempt from the ethics board approval of the university because the data are publicly available and accessible.

Furthermore, the researcher has ensured that the 'Terms of Use' cited on the website at ratemds.com/about/terms should be adhered to.

> "You may read our copyrighted material free of charge, but you are prohibited from the sharing, dissemination, or sale of this material." [4]

1.7 CONTRIBUTIONS

Building on Liu et. al [5] this research study has borrowed a similar methodology in undertaking the analysis on a new dataset for the period of September 2013 to January 2019 and applied LDA, which is a novel application of this method. Liu et al. did not explore the text-mining, which other writers did in USA but not in Canada.

1.8 ORGANIZATION OF THESIS

Chapter 2 – Literature Review: This chapter contains a discussion on the topic of doctors' reviews, Natural Language Processing, Data mining, and Machine Learning. It reviews the different cases where researchers have analyzed doctors' reviews, investigating trends, and other characteristics that are significantly influential on web-based physician ratings. The chapter further explores the text-reviews of the doctors in Canada. Furthermore, the chapter touches on the significant factor that influences the likelihood of healthcare providers' online rating, looking at the extents to which websites for rating doctors are being used currently.

Chapter 3 – Methodology: This chapter contains a discussion on the research methodology. The chapter outlines the steps performed for the data collection, pre-processing, and data analysis.

Chapter 4 – Results: This chapter details the results obtained both for analyzing the meta-data and text-reviews using topic-modeling technique LDA for the doctors from RateMDs.com for the period September 2013 to January 2019.

Chapter 5 – Discussion: This chapter explains the findings following the methodology employed. It explains why the results showed healthcare providers in some of the specialities, such as plastic surgery, family doctors and dentistry, have a higher likelihood of being rated than others such as podiatrists, radiation oncologists, and osteopath. It also shows how the results from this research add more information to work done by previous researchers, citing similarity in the quality of rating for healthcare providers in surgical specialities, obstetrics, gynaecology as well as primary care. The chapter also explores other factors that influence the rating of healthcare providers within the provinces, such as influence of economic prosperity.

Chapter 6 – Conclusion: This chapter concludes this thesis with a summary. It acknowledges the fact that websites such as RateMDs.com provided healthcare consumers with the opportunity to review their respective healthcare providers. Furthermore, the chapter summarizes the new findings to reveal the significant differences in rating according to specialties and regions.

.

CHAPTER 2 LITERATURE REVIEW

2.1 ONLINE DOCTORS REVIEWS

Opinions are a significant part of the human beings, as such, they significantly influence our behaviour when it comes to decision making. The perception and belief of humans regarding reality and the daily-routine choices that we make, inherent significantly from the interpretation of the world by fellow humans [6]. Therefore, this implies that the analysis of people's opinions enable oneself to gain valuable insight on the behaviours of the people expressing the particular opinions as well as the behaviour of those that are influenced by the respective opinions. Having opinions is one of human traits, as evident on the internet, which is why the growth of web 3.0 has seen an increase in the generation of massive user-generated content. The fact that this content, thanks to the same internet technologies, is easily accessible by people that come from different backgrounds and cultures irrespective of the geographical location, implies that there are many opinions that influence different people at the same time. This being the case, more people are encouraged to present their opinions on different platforms freely, and these opinions influence their decisions as well as those of others that agree with these opinions. Websites are the platform that allow social interactions and exchanging reviews amongst fellow beings. This development in web technology has grown into one of today's most popular avenues for people to express their opinions on several issues that affect them [7]. To many consumers, review websites are an opportunity to insightfully evaluate a product and service before making a purchase decision, both for high-value and/or low-value items. The generation of content keeps going on and changing from one form to another as the technology itself advances. The greater proportion of this data are not organized in any form. Irrespective of this factor, however, if well organized and analyzed the information can be of significant use to researchers.

There is an increasing popularity of online platforms on which people can go and search for information about health and health practitioners. Healthcare consumers not only are searching for information about bio-medicine, such as different diseases and ailments, but also about the healthcare facilities, such as hospitals, and the healthcare providers. Empirical research shows that a significant number of the health-related online searches by healthcare consumers are related to healthcare facilities, such as hospitals, and healthcare providers. Those resarches are related to the objective performance measures, i.e. how good the healthcare providers, quantified in terms of reviews and ratings. The healthcare consumers have realized an avenue to give first-hand feedback about healthcare providers and their experiences of the visit to the healthcare facility [8]. Healthcare consumers are encouraged to review, rate, and share candid opinions on their experiences with the respective healthcare provider on popular doctor rating sites in North America such as healthGrades.com, RateMDs.com, Google reviews and Vitals.com on various factors such as healthcare provider's personal demeanour, staff, quality, timing and knowledgeability. Using these website, healthcare consumers might rely on the reviews and ratings to opt for the services of a particular Healthcare provider. The ratings that doctors on these platforms get come in handy to healthcare consumers in an easily accessible and convenient manner when they are making healthcare-related decisions, and they also provide highly reliable insights on what a healthcare consumer's perception of a doctor especially when the raw data is empirically analyzed. Several researchers have studied the concept of deriving or rather defining the sentiments that users have on doctors through comments and reviews.

Researchers have, in many different cases, analyzed the doctor reviews that are given by healthcare consumers on physician-review websites. There are trends in doctor reviews, and Gao et al. [9] researched to investigate the different characteristics and factors that significantly influence web-based physician ratings. Additionally, the majority of the reviews were found to have been positive generally. In research, comparing doctor reviews based on surgeon volume, Segal et al. [10] found through their analysis that there is a significant difference between more frequented and less frequented surgeons based on the text reviews present, numerical ratings, number of reviews that are critical and positive that they each get. According to Liu et. al [5], little information is available regarding the use of Canadian doctor review websites [11].

Therefore, the intention of this research study was to conduct analysis on the meta-data through the adoption, and reliance of healthcare providers' rating websites in Canadian using the data scraped for the period from September 2013 to January 2019.

The investigation was intended for gaining insight on whether the ratings given to different healthcare providers were any different with respect to the areas of specialty, the geo-locations, and also to discover if there were any visible trends worth considering in the data for the period. The study compared the frequency of the ratings that each of the respective healthcare providers had received to their specialty disciplines to identify whether the means of the different groups of healthcare providers were significantly different empirically. Since the ratings were given by Canadian residents, the dataset was representative of Canadian population's opinions [12]. The hypothesis of the study was based on the insight acquired from reading previous studies of the same topic [4, 12].

Furthermore, some studies suggested that there is high possibility that healthcare providers in some specialties such as family medicine get more reviews from healthcare consumers than those in specialties such as radiology and pathology [14].

The understanding of the reviews based on this literature informs that the majority of the ratings and reviews that were received by healthcare providers were positive. The expectation was that a statistically significant difference would be observed in the ratings based on the specialty of the healthcare providers as well as the geographical locations. The data chosen for the research study is scraped from RateMDs.com (a Canadian website) for the period from September 2013 to January 2019. RateMDs.com is widely recognized as one of the biggest Canadian physician-rating websites [5]. The findings of the research showed that there actually are some physician specialties that are more favoured thus more likely to be reviewed compared to others. The primary factors observed during the research, which significantly contributed to disparity in the likelihood of rating included healthcare consumer density by region as well as healthcare consumer expectations [15]. When it came to the ratings that were awarded by healthcare consumers with respect to the quality of service that they received from their healthcare providers, they were the same for surgical processes, medical specialists, gynecology, obstetrics, and the primary care [16]. This was not the same for a different category of doctors that had pathologists, radiologists, and anesthesiologists on the list. Other research works have shown that both pediatricians and surgeons are more favored when it comes to receiving healthcare consumers ratings compared . Regarding the frequency healthcare consumers rated a physician, there is a clear distinction among the specialties.

Healthcare consumers rated specialties like family medicine, obstetrics and dermatology more than radiology and pathology healthcare providers. This statistical probability phenomenon was identified from the data scraped for the period [17]. The information was limited only to the healthcare providers whose data was part of the dataset that was being used.

The circumstances in which the healthcare providers and their healthcare consumers interact was another significant factor that affects the probability or likelihood of a given physician's online rating healthcare consumer. An example is with pregnant women and surgery. The whole process of labour, delivering a baby, it is hard for the healthcare consumers to take note of the quality of the healthcare provider's service since the healthcare consumer is experiencing pain and is asked to strain and push the baby out [18]. In surgery, most of the healthcare consumers are always unconscious and sometimes even artificially dead, so they cannot quite tell how well they were operated on or whether their respective attending healthcare providers were doing a good job based on their personal definition of what a good physician should be like. This makes it hard for both healthcare consumers to review and rate their healthcare providers, which could affect the number of ratings that both surgeons and the healthcare providers receive. Another example of a circumstantial drawback is being a family doctor. Since being a family doctor physician usually guarantees direct healthcare provider - consumer interactions [19]. This, therefore, provides such a physician with the probability of being rated very well and more frequently than a pathologist or radiologist as their healthcare provider - consumer interactions are limited. Another factor that significantly affects how healthcare providers are rated is the fact that healthcare consumers, often, tend to attribute all the care that they receive to a single service provider. For instance, healthcare consumers rarely consider their treatments to have involved the input of more than one specialist whenever their care involve the input of healthcare providers from different specialties [20]. This, therefore, means that they will only rate the ones that they fancy the most if they choose to rate and review the healthcare providers that extended the service to them.

The other significant factor that could affect how healthcare providers were possibly rated by their healthcare consumers was the geographical location of practice.

Since the national accreditation and education standards in Canada are strict, it is difficult to imagine that the quality of a healthcare provider is dependent on the region in which they practice. However, the study's findings show that there is a statistically significant difference in ratings and reviews between healthcare providers based on the geographical locations of practice. Healthcare providers that practiced their disciplines in the Eastern provinces were found to have more reviews and ratings in comparison to those operating in provinces that are to the west of Ontario [5]. A known cause that my study revealed was the reason as to why ratings in a given area might be significantly lower than it is in other regions is the accessibility to healthcare services. Healthcare providers operating in provinces with easy access to healthcare are usually less likely to be rated frequently as compared to those that work in provinces with challenging access to healthcare [21]. Like this realization, research reported lower satisfaction in healthcare consumers that live in more economically prosperous provinces with respect to gross domestic product, such as Ontario, Alberta, and/or British Columbia. This finding contrasts a theory by Gridoroudis [22], which claims that more healthcare consumer satisfaction could be explained by the economic prosperity of a particular region.

According to Galizzi et al. [23], the degree to which doctor-rating websites are used and known was researched with the help of a sample population from London to gain valuable insights on the key predictors that encourage people to make use of such websites that rate healthcare providers. The experiment involved providing the respondents with a questionnaire that was self-administered to conduct an assessment of the determinants as well as the extent of how aware people are about such rating websites, how much the websites are used, as well as the degree of intention to use such websites by Londoners. The conclusion of this experimental research that shows physician-rating websites are significantly informing the healthcare related decisions that people make with respect to who people can consult and so much more. According to Keckley et al. [24], in the United States, , around 47% of people search online for their healthcare providers, 37% look at website rating websites and 7% of those that look up their health practitioners made sure to post a review of them online. In a different study, it was found that 15% of healthcare service consumers make an effort to compare online information about different hospitals before settling on the one they will attend or go to while 30% of healthcare service consumers make this comparison between practitioners before settling on the one to go to. It is therefore very clear that, for the sake of accurate evaluation, there is need to include a measure of sentiments in the reviews that consumers give to healthcare providers with respect to their professional qualities and present them in simple numeric levels that are very easy to comprehend.

2.2 MACHINE LEARNING AND TEXT CLASSIFICATION

Machine Learning (ML) is the use of artificial intelligence (AI) to enable computer systems the learning capability and improve from those learning experiences without the need for explicit programming. Machine learning has enhanced from a simple laboratory interest into a more practical commercial use technology [25]. ML concentrates on developing the computer programs that can not only can retrieve data but also can use it for learning. The ML process usually involves actions such as observations, example data, instructions, and direct experience that helps the computer look for a pattern to make better decisions in the future based on the training dataset. ML algorithms can be grouped by their learning styles and similarity of their function [26]. The Algorithms based on the learning-styles are;

(i) Supervised
(ii) Unsupervised
(iii) Semi-supervised

while, the algorithms based on the similarity of their function are;

(i) Regression
(ii) Instance-based Algorithms
(iii) Regularization Algorithms
(iv) Decision-Tree Algorithms
(v) Bayesian
(vi) Clustering
(vii) Association Rule Learning
(viii) Artificial Neural Networks
(ix) Deep Learning
(x) Dimensionality Reduction Algorithms
(xi) Ensemble

Furthermore, the employment of machine learning algorithms in text classification and analysis is common place. Different machine learning algorithms are used for this purpose. In their work, Kennedy et al. [27] made use of a random forest classifier algorithm to depict whether there were any forms of harassment in both Reddit, Twitter, and The Guardian posts. The presentation of the posts being analyzed was done based on different characteristics such as term frequency-inverse document, frequency of bigrams, short character sequences and unigrams; sentiment polarity; the source from which they were gathered; and hashtag and URL token counts. In research that classified twitter 'hates speech' posts, Gambäck and Sikdar [28] employed Convolutional Neural Network (CNN). The employed CNN model was applied along with several different embedded features that included word vectors and random values. Different classification techniques are employed in sentiment analysis to help with the identification of where a given text lies sentimentally. The techniques help the researcher with insight into whether the text is positive, neutral, or negative in nature. Semantic analysis can be conducted on very large datasets, and the results of the analysis can be applied in different contexts. A good example is when a given company analyses all the feedback and reviews that they get from their customers with respect to a given product that they may have introduced into the market. This way, they will be able to know whether the product is performing according to the performance projections that they had or not and thus make an informed decision on how to deal with the concern. Another application area could be in politics where a given politician looks to find out whether the messages that they are sending with respect to their campaigns are resonating with the voters. Physician reviewing websites also provide so much information that can be analyzed through text classification and sentiment analysis.

Recent machine learning developments have shown that the CNN model can yield text classifiers with high fidelity. CNN can also obtain significantly high-performance rates when applied in Natural Language Processing (NLP) tasks, even in sentiment analysis. One of the most notable works is the application of trained CNN on top of word vectors that are pre-trained in classification tasks at the sentence level and providing accurate results in generic analysis of sentiments. Engineering of features can be replaced by CNNs that display the abilities of this kind of models that capture high-level features, thus leading to text classifiers that are even more superior. The research attempted to rate healthcare providers in terms of healthcare consumer reviews [29]. The qualitative analysis that they conducted involved statistically analyzing 712 reviews that were sourced from two rating websites. The sample reviews were of 445 doctors that provided primary care that belonged to four different urban locations in the United States. The observation made by the research was that the majority of primary care healthcare providers' reviews are positive. According to their outcome, what healthcare consumers refer to as care is not only based on the relationship and experiences that they share with the healthcare providers but also with the staff, how much access they have at the health facility and last but not least, the convenience that both the healthcare provider and the facility provide the healthcare consumers with. This research represents the traditional way of data analysis, which is very different from the employment of CNNs in an analysis. Ganu et al. [30], conducted a research that involved text classification that was sentiment based that purposed to identify the sentimental and topical information from restaurant free-form text reviews to be able to enhance user experience in how they access reviews. A more improved version of the CNN called the Dynamic Convolutional Neural Network (DCNN) was adopted by Kalchbrenner et al. [29] for sentence semantic modelling. Different from the CNN, the network makes use of Dynamic k-Max pooling, which is a global pulling operation over the linear sequence. The sentences that it analyses are of varying lengths and in analyzing them it induces a feature graph capable of capturing quite explicitly both the long and the short-range relations over the respective sentence. This model is highly effective given that it performs well when applied to a varying range of NLP tasks, sentiment analysis included, and reports an error reduction that is above 25% when applied in Twitter sentiment analysis. It has been shown by Zhang et al. [31] that through the use of convolutional networks, deep learning can be applied in text comprehension down from the character level up to concepts of the abstract text. These researchers go on to show further how Convolutional Networks do not need any knowledge on either the semantic or syntactic language structure for them to give reliable text to understand benchmarks, especially in comparison to earlier approaches that required the starting point of having a dictionary of words and hard-wired into the models is structured parsing.

A series of experiments were conducted by Yoon Kim [32] with convolutional neural networks (CNN) on top of word vectors that had already been trained to classification tasks of the sentence level. The findings of this research are that CNNs achieve excellent results in many different areas when used together with static vectors and when some hyperparameter tuning is conducted. Proposed by the same research is an architectural advancement that would allow for both static vectors and task-specific vectors in the CNNs. The research mainly focuses on comparing multiple convolutional neural network variants, which included non-static, static, random, and multichannel based on the training that the vectors had been put through. A sentence is viewed as a concatenation of words and filters are applied to every window of words available so that a feature map is produced. Once this feature map is created, it required the application of max-overtime pooling operation to highlight which the most significant features is in each of the feature maps. The vectors employed have one channel that is from static training and another one that is enhanced through backpropagation. Many natural language standard processing problems are dealt away with this approach and among them is question classification and sentiment analysis.

Sentiment analysis in some research works has also been conducted using dependency pattern/trees. In research by Agarwal et al. [33], they employed hand-crafted rules in the extraction of sentiments or rather semantic information from sentences. This information was combined with the semantic information found in the Massachusetts Institute of Technology Media Lab ConceptNet Ontology [34]. The concepts that were extracted beforehand were employed in training a machine model intended for learning pattern in the text, which was made use of in the categorization of the respective documents into negative and positive categories. Although the employed dependency pattern in this research only contained two words, there is a possibility of including more words in both indirect and direct relations. In research by Wawer [35] they induced the respective dependency pattern using target-sentiment (T-S) pairs and recording the paths between the T and S words into the dependency tree, with respect to the sentences in the corpus that was supposed to be analyzed. To help with the targeting of opinion words, the patterns were supplemented with conditional random fields.

2.2.1 Types of Machine Learning Algorithms

This section details two types of Machine learning algorithms i.e. supervised and un-supervised learning algorithms;

A. **Supervised Learning**: Supervised learning algorithms involve the application of an algorithm to gather the insight on the historical data to predict the future. It mainly involves the assessment of an established dataset, with the learning algorithms producing an incidental function that goes ahead to predict the output values. It allows the system to deliver targets for any modern input and find errors that can then be used to transform the model further.

B. **Unsupervised Learning**: Unsupervised learning algorithms are regarded as the true artificial intelligence and are usually applied whenever the information being used to train the system is neither labelled nor classified [36]. The algorithm usually studies how a system can infer an action without human intervention, which can then be used to define a hidden construct from data that is not labelled. Although the system might not figure out the correct output, by exploring the present data, it can describe the hidden structures from the data.

This research study have utilized the un-supervised learning algorithm for topic-modeling i.e. LDA.

2.3 WEB SCRAPING

Web scraping, also known as Web harvesting, Web data extraction, or Web automation [37] is a technique to extract mostly unstructured data from HTML tags of a web page. Web scrapers, also known as web robot, which imitates the communication between the Web servers and the humans in a conventional Web traversal [38]. Web scrapping is like Web indexing, an information retrieval technique used to index information on the Web through bot. The scraped data focuses on the transformation of unstructured data on the Web that can be stored and analyzed in a central local database or spreadsheet. Some of the applications of Web scraping are; online price comparison, weather data monitoring, website change detection, Web research, Web mashup, and Web data integration [37].

CHAPTER 3 METHODOLOGY

The study collected data from RateMDs.com on the Canadian healthcare providers. RateMDs allow the healthcare providers to create their basic profile for free, while it does not require the healthcare consumers to have registered to use their services. It is currently one of the most widely used websites for rating healthcare providers in Canada [39].

Furthermore, the website uses a scale of between 1 and 5, where 1 stands for terrible, 2 stands for poor, 3 stands for okay, 4 stands for good and 5 stands for excellent. The rating, encompasses a wide domain, including aspects of punctuality, knowledge, among other aspects that define the performance of healthcare providers [40]. For each physician, the profile is usually created to allow for easy search by raters.

3.1 STUDY TOOLS

The following software tools and OS platform were used for this thesis.

Operating System Ubuntu 18.04 x64

Data Analysis Python 3.7.4

Python Packages and Libraries BeautifulSoup, re, NLTK, spaCy, Gensim, Numpy, Pandas.

Document Preparation MS Word and MS Excel.

Figures MS Office 365 Word, Excel and Power Point.

Bibliography Management Mendeley Desktop 1.19.4.

Terminal MobaXterm Home Edition V11.1

Tunneling MobaXterm Home Edition V11.1

File Transfer WinScp 5.15.3

Cloud Platform Digital Ocean (16 GB Memory; 320 GB Disk; TORI Ubuntu 18.04 x64)

3.2 DATA COLLECTION

This section details the steps taken for collecting (web-scraping), pre-processing, and cleaning the dataset used in the thesis.

(1) Webpage for each doctor: Each registered doctor on RateMDs.com has been assigned an individual webpage, which is accessible via a URL. There is a total of 169,983 webpages scraped, containing the meta-data as explained in step 2, from the website for the period from September 2013 to January 2019.

(2) Meta-data of the doctors in Canada: As shown in Figure 1, a bot was developed in Python using BeautifulSoup, TextBlob, re, Pandas, to go through each doctor's webpage and fetch the following meta-data:

Dr's Name, Gender, Practice, Total No. of Reviews, Average Rating, Province, City.

The meta-data was instrumental in getting the descriptive statistics for comparing physician specialties, spread of Gender, and ratings among the provinces in Canada.

3.3 ANALYZING THE DATA SCRAPED FROM RATEMDS.COM

Following the data collection; steps for analyzing, model development and deployment were performed based on the data from RateMDs.com (Canada).

3.3.1 Steps Performed to analyze the dataset

There was a total of 169,983 URLs (1 Doctor = 1 URL); of which the meta-data for 132,479 doctors were scraped as the rest of the URLs were bad. Then, the review texts for 84,292 doctors have been collected (the rest of the doctor's reviews were not available). A doctor could have one or multiple pages of reviews available.

For further comprehension, let us take a case scenario of Dr M'. Figure 1 is a screenshot of the main page of RateMDs.com.

Figure 1: The main page of RateMDs.com containing the web links for all doctors.

As an example, Dr. Ms' web page [41] is displayed following a click on his name as shown in Figure 2.

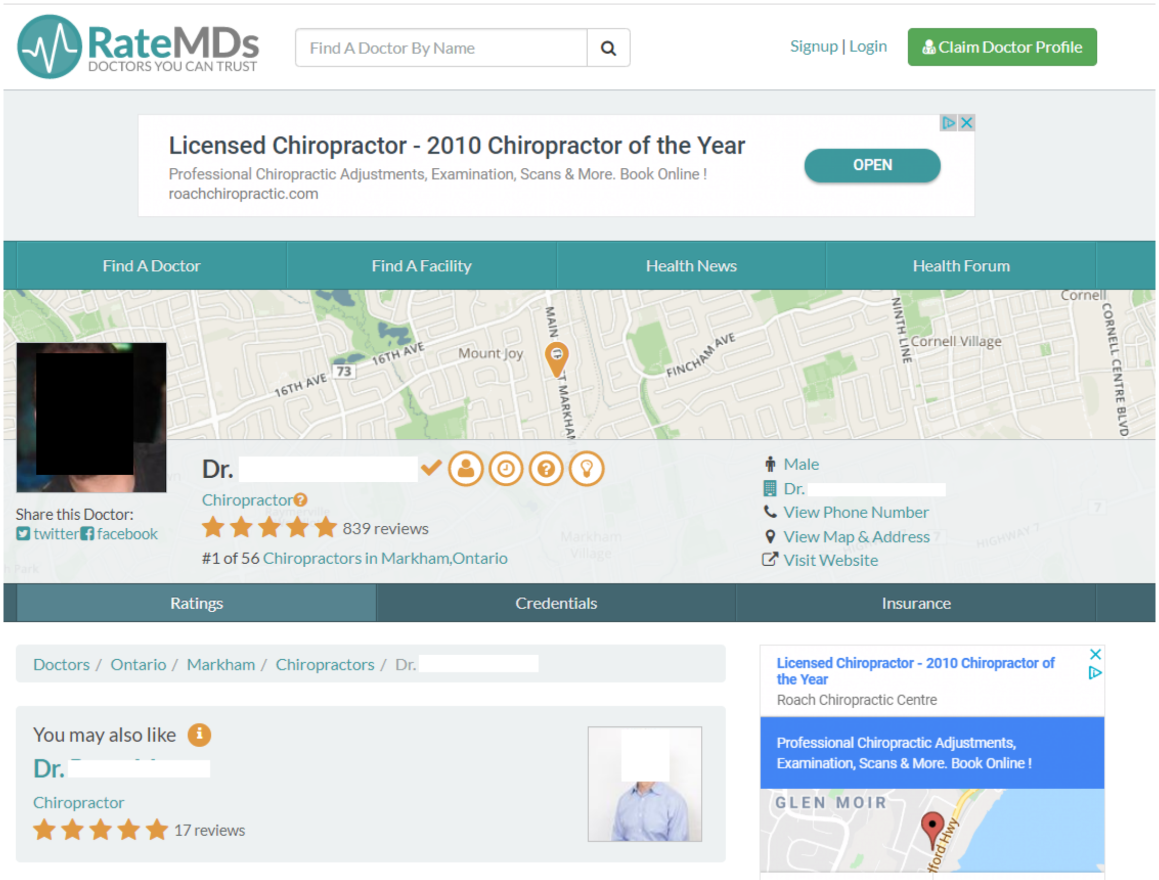

Figure 2: Dr Ms' webpage showing all the reviews submitted by his customers.

There is a total of 85 pages containing reviews for Dr M's. Page 1 to 84 contain 10 reviews per page, while page 85 has only nine reviews. Hence, this shows that every doctor has a different number of review pages with differing number of reviews.

Filtering the data by doctor's specialty using MS Excel: Once the data has been scraped, it was pre-processed for further analysis. The data has been filtered, using MS Excel, according to the specialties, i.e. Cardiologists, Dentists, Family Doctors, Psychologists etc.

Collecting text reviews: Following data filtering in Step 1, a Python code was executed to collect the text reviews for the specialties mentioned in the previous step.

The following Python libraries and packages are used for analyzing and visualizing the data:

(i) Numpy, (ii) Pandas, (iii) NLTK, (iv) spaCy, (v) Gensim, (vi) Sklearn

The following sklearn classes are used;

(i) LDA, (ii) Counter Vectors, (iii) Grid Search, (iv) Cross Validation

The following libraries for visualization are used;

(i) pyLDAvis, (ii) Matplotlib

The following steps are performed to run the LDA (Latent Dirichlet Allocation) model (please refer to Appendix B for a detailed execution of the model along with its results);

(i) Read the data to ensure each row has a review

(ii) Convert the data frame to a list

(iii) Used RegEx to remove quotations, multi-lines

(iv) Lemmatization – spaCy has been used

(v) Vectors has been used to count the frequencies of occurrence of words

(vi) Build LDA model

(vii) Create a search param (2x)

(viii) Apply Cross-validation as 3 folds (default)

(ix) Calculate the Learning decay

(x) Plot the grid using matplotlib.

3.4 A COMPARISON AMONG CANADIAN PROVINCES

This research study has borrowed a similar methodology as Liu et al. [5] and undertaken a similar analysis on a new dataset for the period of September 2013 to January 2019 and applied LDA, which is a novel application of this method in Canada. Moreover, the studies involved calculations of the average number of ratings and mean rating scores for each physician as well as each province. In a bid to compare the relative proportions of the healthcare providers, they were grouped according to the specialties, following the grouping by Canadian Institute of Health Information (CIHI).

Statistics: For the statistical analysis, the study aimed at examining and comparing differences between ratings of healthcare providers by their specialty and geographical location. The physician ratings were further categorized into either favorable or unfavorable, with the reference point being the median of the average rating, computed as 3.68 [38]. We then constructed a binary variable to indicate whether the physician received a favorable or unfavorable rating. The ones who received an average rating of less than 2.87 were considered to have unfavorable rating while those that received a rating higher than 2.87 were considered to have been rated favorably [39].

3.5 LATENT DIRICHLET ALLOCATION (LDA)

For the second part of the study i.e. applying LDA to the text-reviews, where a similar methodology has been adopted from [13] which was conducted in the USA.

Latent Dirichlet Allocation (LDA) is a generative text model. The model works in such a way that words within a document replicate mixed latent topics. Each of the tokens is usually linked with variables of the latent topics. The factorial LDA—often denoted by f-LDA—usually generalizes LDA to allow for the association of each token to the latent of variables [39]. For instance, in considering a two-dimensional model, every token can often be related to two variables that correspond to the sentiment and the topic.

As our method of determining the common topics that are discussed in medical reviews, LDA was chosen since it is based on a probabilistic generative model [42]. The method is fully unsupervised and is effective in identifying common discussion topics from a collection of documents. It involves the automatic identification of topics without necessarily requiring manual annotation or prior knowledge [14]. The consideration made it particularly attractive for the task at hand as the study aimed to discover the topics that were commonly discussed while reviewing health providers instead of developing hypotheses about important elements of the practice of health providers to the consumers and authenticating them using analysis of data.

As a probabilistically generative text collection model, LDA entailed the representation of documents as a random mix over latent topic. For each topic, the characterization was done by a distribution over words [43]. The topic distribution was further taken to be from a set of parameterized Dirichlet distributions, meaning that the words within a document were consecutively generated by repeated sampling of a topic as per the distribution topic before selecting a word following the chosen topic [43].

The model's main utility was chosen by reversing it before inferring the unknown topics, and their associated words, from a collection of pre-existing documents [39]. The process of inference involved the calculation of the most likely distribution of topics and assignment of words from observed data. The result was an output of the data was a list of topics, with the likelihood of every word appearing in the topic mentioned in the data [13]. The model did not provide any labels or names for the topics since it was unsupervised. The idea of each topic's subject was mainly obtained by examining the words that were most probable for each topic. These probable words were then assigned labels. The generative nature of LDA allowed it to handle documents that were newly examined, and which did not precisely conform to distributions that were previously observed [43]. Some authors have since compared LDA to other models, which have been mentioned to perform less over-fitting than the others, reporting improved results on tasks of text classification and modeling of documents. For a while now, LDA has been used in numerous tasks including resolution of entities, retrieval of information and image processing, among other tasks [13]. The discoveries have led to the development of several other efficient methods for deductions with LDA. This study employed a standard implementation of the model, using Gibbs sampling for inference and estimation of parameters.

For the study, it was important first to describe the main method of computation that would then be relied on. We used the Latent Dirichlet Allocation (LDA) for the identification of salient aspects in the review of health-care providers. The computation method was then customized to answer the research questions by addressing two major challenges in particular. The first challenge involved the selection of dataset and the processing unit on which the LDA was applied [13]. The second challenge involved the determination of an optimal number of aspects that were discussed in the reviews, otherwise known as the model order.

3.5.1 Data

The creation of datasets was done using a collection of procedures that involved collecting a corpus of reviews from RateMDs website [44]. In the pre-processing step, the HTML pages were used to extract portions containing reviews as well as the designation of specialty for every provider. The reviews were then separated and tokenized to specific sentences, followed by the removal of stop words [45]. The dataset containing reviews was then stratified into sets of reviews such as general practitioners, psychiatrists, dentists, and gynecologists, among other strata.

3.5.2 Prerequisites

As a prerequisite, the process involved downloading NLTK (a Python library) stopwords and the spaCy models. As mentioned earlier, the model needed stopwords from NLTK and spaCy's model for the pre-processing before involving the spaCy model for lemmatization [43].

3.5.3 Importing Packages

The process used re, genism, spacy, and pyLDAvis as the core packages for the study. Data handling and visualization for the study used pandas, numpy, and matplotlib [44].

After preparing the stop words, the process involved importing text reviews data. This step is broken down into two categories, i.e. specialties and geo-location. Following specialties are chosen; Acupuncturists, Cardiologists, and Dentists, datasets [13]. My rationale was to select a balanced dataset to be analyzed. While, all the Canadian provinces i.e. ON, AB, MB, SAK, NS, NB, NL, NT, NU, and BC have been chosen for the analysis.

The next step in the process involved the removal of emails, newline characters, and extra spaces, which were considered distracting. The removal involved using a regular expression. Afterwards, the tokenization technique and clean-up texts was applied [43]. Furthermore, the punctuations and other unnecessary characters were removed, forming each sentence into a list of words.

3.5.4 Creating Bigram and Trigram Models

The next process in the model involved creating bigrams—two words that frequently occur together—and trigrams—three words that frequently occur together [46]. The building and implementation of the bigrams and trigrams used Gensim phrases model. Once the bigrams model was ready, the following step involved defining functions to remove stop words followed by lemmatization and then calling the bigrams sequentially. The step led to the creation of the dictionary and corpus needed for modeling of the topics. As the requisites to train the LDA model was obtained, the number of topics was supposed to be provided in addition to the corpus and the dictionary as a way of building the topic model.

The LDA model in use was built with 5, 7, 9 different topics, with each topic being a combination of keywords and the keywords contributing to the weight of the topic (please refer to Appendix B for code snippets). The weights represented the importance of each keyword to the specific topic [46]. The perplexity of the model and the coherence of the topic provided a convenient measure to be used in judging the quality of a given topic model.

3.5.5 Visualization of Topics-Keywords

With the LDA model already built, it was important to examine the topics that were produced as well as the associated keywords using the interactive chart of pyLDAvis, working better with jupyter notebooks [13]. For a better quality of the topics, however, the Mallet's version of the LDA model needed to be used [44]. In finding the optimal number of topics for LDA, the approach involved building many models with different values of the various number of topics [43]. The model involved finding the dominant topic in each sentence; determining the topic on which a given document is about one of the practical applications of the modelling. Before the distribution of topics across the documents, it was important to find the document that was most representative for each topic [44]. The documents were essential in assisting with making enough sense of what each topic is about. Finally, the distribution of topics across the document was to help with understanding the volume and distribution of topics to judge the breadth of its discussion. Please refer to Appendix B.

3.5.6 LDA for Reviewing Health-Care Providers

The LDA model turned out to be more efficient at finding aspects of health-care reviews that were rateable. The local version of LDA was seen to be able to find rateable aspects, which operated on individual sentences instead of documents, in various domains, including physician reviews. The aspects did not, however, require additional information [47].

3.5.7 Model Order

The aspect of the order of the model was essential for unsupervised learning as it aimed at determining the correct cluster numbers. The common approach involved relying on a procedure involving validation of clusters.

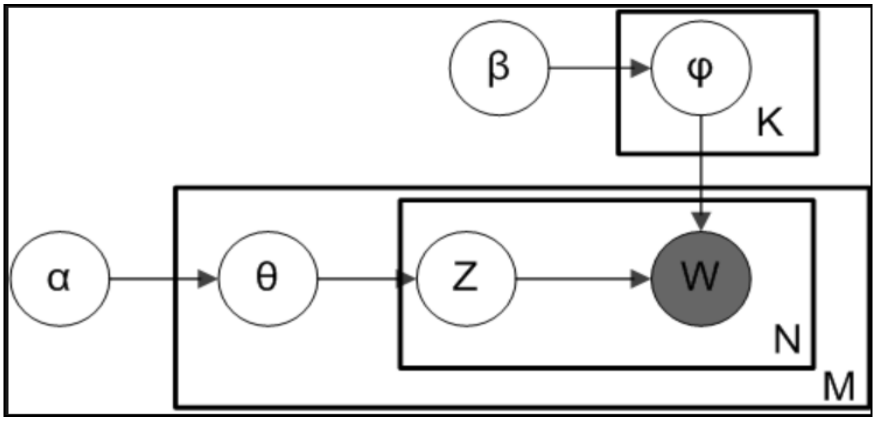

Figure 3: LDA graphical representation, with the boxes as plates that represent replicates. The inner plate represents the repeated topics and word choices in a document while the outer plate stands for the documents. Adapted from Jordan and Mitchell (2015)

Figure 3 shows a plate model of LDA. It entails the comparison of different model orders, choosing the model that had the most consistent clustering [29]. For the validation process, a cluster was required for each corresponding aspect, and each sentence was labeled to be appropriate for the cluster that had the most likely aspects.

Following the sentence assembly in the data, denoted as D, and the connectivity matrices, denoted as C and C', the consistency function, denoted by F was defined. For the connectivity matrix, cells i and j were equated to 1 if the sentences di and dj belonged to the same cluster. The consistency function was denoted as follows:

$$F(CC') = \frac{\sum_{i,j} 1\{C_{i,j} = C'_{i,j} = 1, d_i, d_j \in D'\}}{\sum_{i,j} 1\{C_{i,j} = 1, d_{i,} d_j \in D'\}}$$

Figure 4: The consistency function

The procedure that followed is as described below:

As shown in Figure 4, running the LDA model with k topics on the data (D) to get the connectivity matric (C_k). The second step involved creating a comparison connectivity matrix (R_k) based on the random assignments of each instance that was drawn uniformly. The random subset D^i from D of size $\delta|D|$ was sampled [11]. The LDA model was then run on D^i to get the connectivity matrix C_k^i before creating the comparison matrix R_k^i based on the random assignment of the instances in D^i that were drawn uniformly. The $score_i$ (k) from subtracting $F(R_k^i, R_k)$ from $F(C_k^i, C_k)$. The average score was then returned over q iterations.

CHAPTER 4 RESULTS

Table 1 shows that there was a total of 414,442 ratings for 44,530 healthcare providers for the period from September 2013 to January 2019. RateMDs.com uses a scale of 1 to 5. The ratings were relatively positive, with the overall average rating of 4.8 and a standard deviation of 0.1. The data represented an average of 8 ratings per physician. The specialization with the highest number of rated healthcare providers was for family doctors, which had a total of 13,614 unique healthcare providers that were rated, with an overall mean rating of 4.68 and a standard deviation of 0.25. The category that had the second highest number of rated healthcare providers was for dentists followed by chiropractors, with overall mean ratings of 4.79 and 4.85 respectively the standard deviations were 0.24 and 0.20 for the two categories respectively. With the family doctor category having the highest number of rated healthcare providers, it translated to the highest number of ratings as well, with a total of 139,084 ratings.

Various categories, however, had very few ratings, with the likes of urogynecologist, psychotherapist and podiatrist categories having only one rating at the time of the study.

4.1 DIFFERENCE IN QUALITY OF RATING BY SPECIALTY

The study further assessed the difference in the quality of ratings of the doctors by their specialization, as depicted in Figure 6. The results showed that some of the specialties were more likely to be rated higher than others. Among the practices that were more likely to be rated above the 50th percentile included Acupuncturist, Family Doctors, Chiropractors, and Dermatologists, among others. On the other hand, some of the categories whose ratings were below the 50th percentile included pain management specialists, gastroenterologists, and the oral surgeon. The probability of favourable rating was calculated with respect to the median, which was calculated as 3.68.

Table 1: Number of ratings, unique rated healthcare providers, rating per physician and overall rating of all physician rated by practice

Practice	ratings, n	unique rated healthcare providers	rating per physician	overall rating	overall rating (sd)
Acupuncturist	4012	913	4.39	4.93	0.16
Addiction Medicine Specialist	244	41	5.95	4.78	0.25
Allergist / Immunologist	894	106	8.43	4.70	0.25
Anesthesiologist	1793	508	3.53	4.88	0.20
Audiologist	34	20	1.70	4.95	0.13
Bariatric / Weight Loss Specialist	335	33	10.15	4.84	0.21
Cardiologist	6422	934	6.88	4.77	0.24
Cardiothoracic Surgeon	1725	155	11.13	4.79	0.25
Chiropractor	27089	3268	8.29	4.85	0.20
Colorectal Surgeon / Proctologist	535	42	12.74	4.89	0.18
Counsellor	1	1	1.00	5.00	0.00
Dentist	74342	7802	9.53	4.79	0.24
Dermatologist	2373	206	11.52	4.70	0.26
Doctor of Naturopathic Medicine	10	1	10.00	4.80	0.00
Ear Nose and Throat Doctor (ENT)	3524	314	11.22	4.71	0.26
Emergency Room Doctor	4767	799	5.97	4.78	0.24
Endocrinologist	2344	246	9.53	4.68	0.25
Endodontist	2897	98	29.56	4.77	0.25
Family Doctor / G.P.	139084	13614	10.22	4.68	0.25
Gastroenterologist	3741	417	8.97	4.71	0.26
General Surgeon	12076	1160	10.41	4.72	0.25
Geneticist	182	27	6.74	4.84	0.17
Gynecologist (OBGYN)	12444	1007	12.36	4.70	0.26
Homeopath	139	12	11.58	4.89	0.20
Infectious Disease Specialist	689	153	4.50	4.81	0.24
Internist / Geriatrician	4316	769	5.61	4.79	0.24
Massage Therapist	67	36	1.86	4.98	0.07
Midwife	1309	272	4.81	4.85	0.23
Naturopath	4814	801	6.01	4.85	0.21
Nephrologist	1800	303	5.94	4.80	0.22
Neurologist	3035	412	7.37	4.73	0.24
Neurosurgeon	2367	207	11.43	4.76	0.26

Practice	ratings, n	unique rated healthcare providers	rating per physician	overall rating	overall rating (sd)
Nurse Practitioner	598	157	3.81	4.85	0.21
Occupational Therapist	117	23	5.09	4.87	0.18
Oncologist / Hematologist	5071	790	6.42	4.77	0.23
Ophthalmologist	5991	636	9.42	4.72	0.25
Optometrist	9915	1431	6.93	4.83	0.21
Oral Surgeon	3756	205	18.32	4.74	0.27
Orthodontist	6905	345	20.01	4.78	0.25
Orthopedic Surgeon	8884	773	11.49	4.72	0.26
Osteopath	4	1	4.00	5.00	0.00
Pain Management Specialist / Physical Therapist	1585	258	6.14	4.80	0.23
Pathologist	69	20	3.45	4.89	0.21
Pediatrician	8740	918	9.52	4.72	0.25
Periodontist	3465	152	22.80	4.80	0.23
Physical Medicine and Rehab / Physiatry	5	1	5.00	4.45	0.00
Physiotherapist	111	2	55.50	4.93	0.10
Plastic / Cosmetic Surgeon, Physician	14287	414	34.51	4.79	0.24
Podiatrist	4185	298	14.04	4.78	0.24
Podiatrist (Serving Burlington & Brampton)	1	1	1.00	5.00	0.00
Psychiatrist	6823	1137	6.00	4.75	0.25
Psychologist	4062	969	4.19	4.85	0.22
Psychotherapist & Counselor	1	1	1.00	4.50	0.00
Pulmonologist	1837	273	6.73	4.78	0.24
Radiation Oncologist	4	3	1.33	4.96	0.07
Radiologist	1086	314	3.46	4.87	0.19
Reproductive Endocrinologist	32	23	1.39	4.76	0.23
Rheumatologist	2273	197	11.54	4.67	0.26
Sleep Doctor	28	23	1.22	4.90	0.17
Sports Medicine Physician	95	21	4.52	4.74	0.24
Therapist	230	30	7.67	4.83	0.22
Urogynecologist	1	1	1.00	5.00	0.00
Urologist	4146	335	12.38	4.74	0.26
Vanity Testing	8	1	8.00	4.84	0.00
Vascular Surgeon /	723	100	7.23	4.77	0.23

Phlebologist

Figure 6 shows the proportions of mean rating by practice categories, with most categories having mean ratings close to the median value of 3.68.

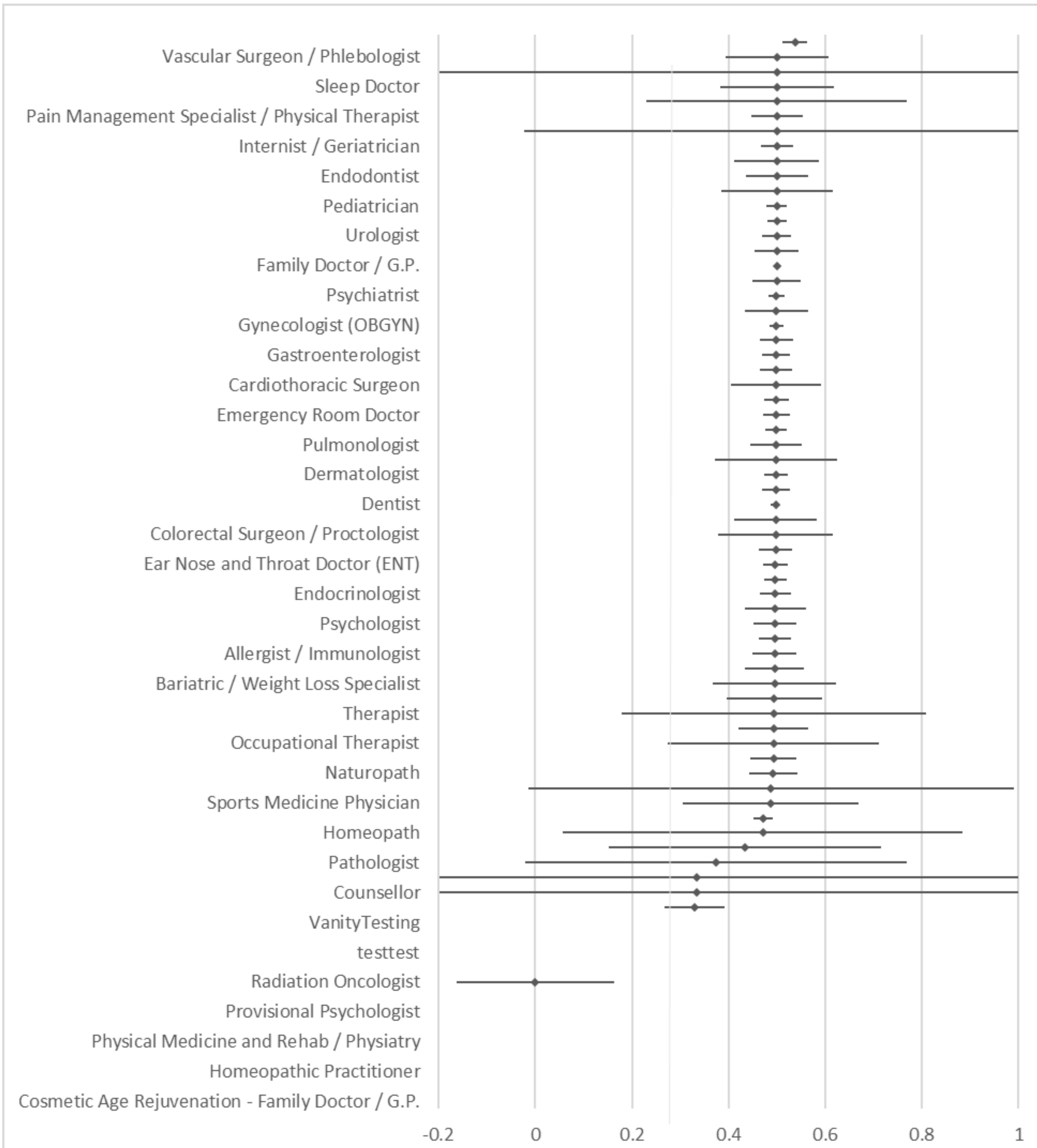

Figure 5: The proportion of mean ratings by practice, in the top 50th percentile of all rated healthcare providers, with a 95% confidence interval for each proportion.

The blank space above the x-axis are all due to the data. The blank space above the x-axis represents those specialties whose proportion of mean ratings by specialty or province, in the top 50th percentile, was zero. The longer error bars tell us that the uncertainty in that data point is high.

4.2 DIFFERENCE IN FREQUENCY OF PHYSICIAN-RATING BY PROVINCE

The study found that the province with the highest number of healthcare providers rated was Ontario, with 49,634 healthcare providers, followed by Quebec and British Columbia, with 30,012 and 18,749 healthcare providers, respectively. The province with the least number of healthcare providers was the Northwest Territories, which had only 135 healthcare providers rated.

As shown in Table 2, Ontario had the highest number of rated healthcare providers, it also had the highest number of ratings, with a total of 545,791 number of ratings. British Columbia, however, came second with 213,302 overall ratings, while Quebec had a total of 197,615 ratings. Northwest Territories turned out to also have the least number of physician ratings, with a total of only 587 ratings.

Table 2: Number of ratings, number of healthcare providers, mean ratings per physician and mean overall rating of all healthcare providers rated by Province

Province	ratings, n	Healthcare providers	Ratings per physician	mean overall rating	overall rating (sd)
Alberta	160045	12473	12.83	3.15	1.80
British Columbia	211172	18188	11.61	2.96	1.87
Manitoba	55138	4548	12.12	3.01	1.84
New Brunswick	25750	2317	11.11	3.26	1.79
Newfoundland	12550	1309	9.59	3.36	1.70
Nova Scotia	38922	3503	11.11	3.17	1.81
Northwest Territories	667	169	3.95	2.38	2.19
Ontario	533421	45312	11.77	2.87	1.92
Prince Edward Island	4342	432	10.05	2.91	1.92
Quebec	5380	414	13.00	4.43	0.53
Saskatchewan	46793	3632	12.88	3.20	1.75

Although Ontario, Quebec and British Columbia provinces had high number of rated healthcare providers as well as overall physician ratings, Newfoundland was the province with the highest mean of 3.36 overall rating and a standard deviation of 1.71. Northwest Territories ended up with the lowest mean overall rating for their healthcare providers as well, with a mean of 2.43 and a standard deviation of 2.18 out of 4.35.

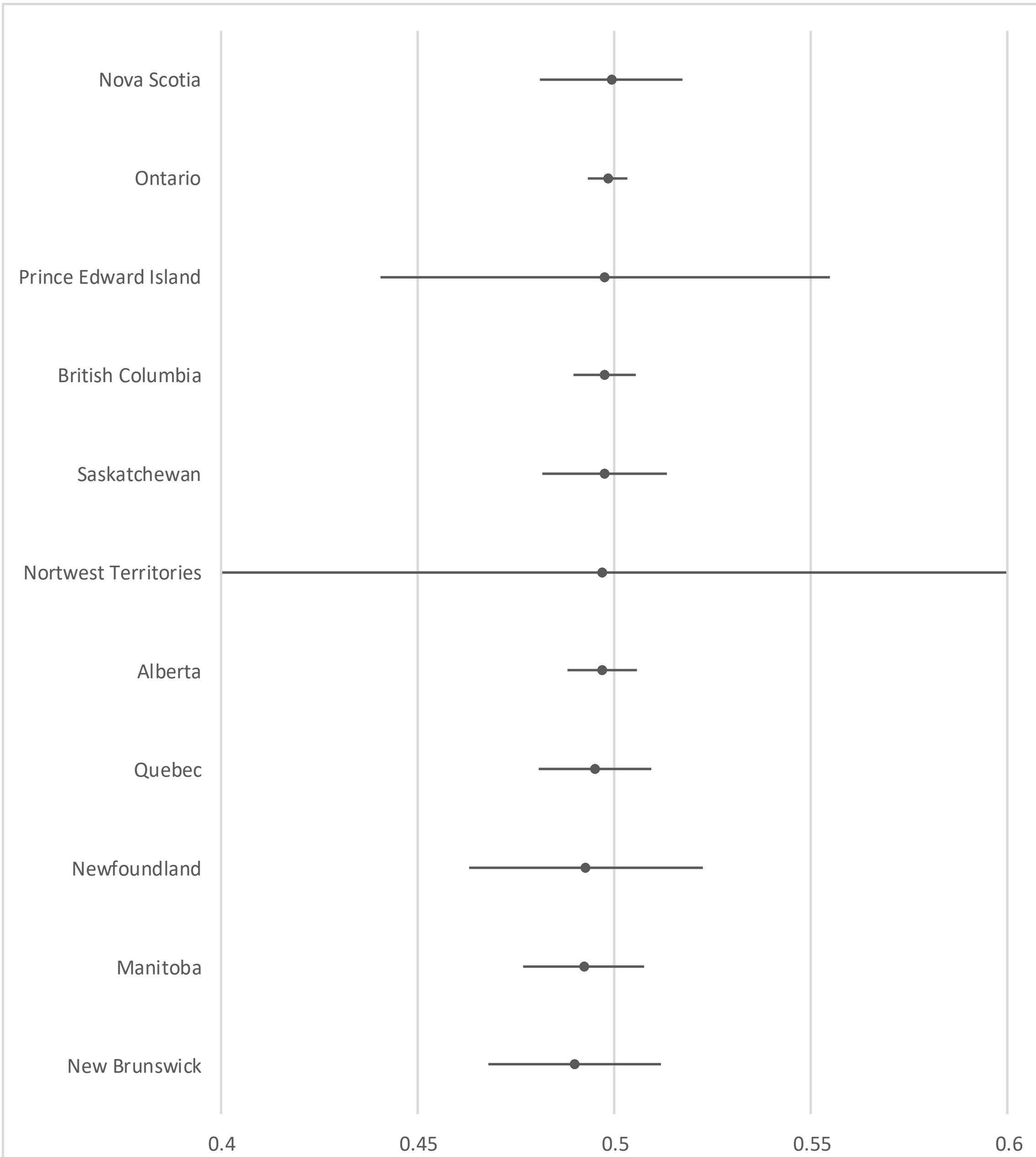

Figure 6: Proportion of Mean Ratings by province, in the top 50th percentile of all rated healthcare providers with a 95% confidence interval depicted for each proportion.

Figure 6 shows that most of the provinces has the ratings per physician ranged between 10 and 12 ratings. The province with the highest rating per physician was

Saskatchewan, with 12.66 ratings while the province with the lowest rating per physician was Northwest Territories, with only 4.35. Other provinces that had relatively high ratings per physician were Alberta (12.26), Manitoba (11.76), British Columbia (11.35), Ontario (11.00) and New Brunswick (10.97).

4.3 DIFFERENCES IN QUALITY OF RATINGS FOR PHYSICIAN PRACTICE LOCATION (BY PROVINCE)

A One-Sample t-test has been applied. The median rating for all the provinces was 3.74. The quality of ratings in all provinces will be compared against this value.

Alberta (AB) - The percentage of ratings above 3.74 was 53.26%. As such, physicians practicing medicine in Alberta were more likely to be rated greater than 3.74 (53.26%, $P<0.001$).

British Columbia (BC) - The percentage of ratings above 3.74 was 49.40%. As such, physicians practicing medicine in British Columbia were likely to be rated less than 3.74 (49.40%, $P<0.001$).

Manitoba (MB) - The percentage of ratings above 3.74 was 50.12%. As such, physicians practicing medicine in Manitoba were likely to be rated greater than 3.74 (50.12%, $P<0.001$).

New Brunswick (NB) - The percentage of ratings above 3.74 was 57.66%. As such, physicians practicing medicine in New Brunswick were more likely to be rated greater than 3.74 (57.66%, $P<0.001$).

Newfoundland and Labrador (NFL) - The percentage of ratings above 3.74 was 58.37%. As such, physicians practicing medicine in Newfoundland were more likely to be rated greater than 3.74 (58.37%, $P<0.001$).

Nova Scotia (NS) - The percentage of ratings above 3.74 was 55.01%. As such, physicians practicing medicine in Nova Scotia were more likely to be rated greater than 3.74 (55.01%, $P<0.001$).

Northwest Territories, Yukon, and Nunavut - The percentage of ratings above 3.74 was 41.42%. As such, physicians practicing medicine in Northwest Territories, Yukon, and Nunavut were more likely to be rated less than 3.74 (41.42%, $P<0.001$).

Ontario (ON) - The percentage of ratings above 3.74 was 47.46%. As such, physicians practicing medicine in Ontario were more likely to be rated less than 3.74 (47.46%, $P<0.001$).

Prince Edward Island (PEI) - The percentage of ratings above 3.74 was 48.84%. As such, physicians practicing medicine in Prince Edward Island were more likely to be rated less than 3.74 (48.84%, $P<0.001$).

Quebec (QC) - The percentage of ratings above 3.74 was 86.71%. As such, physicians practicing medicine in Quebec were highly likely to be rated greater than 3.74 (86.71%, $P<0.001$).

Saskatchewan (SAK) - The percentage of ratings above 3.74 was 53.17%. As such, physicians practicing medicine in Saskatchewan were more likely to be rated greater than 3.74 (53.17%, $P<0.001$).

Please refer to Appendix D for SPSS workings.

4.4 ANALYSIS OF TOPIC-MODELING APPLIED TO RATEMDS.COM

4.4.1 Priors from User Ratings

The reviews from RateMDs corpus contain user ratings for four categories i.e. Staff, Punctuality, Helpfulness, and Knowledgeability. The ratings are in integer (number) form ranging from 1 to 5 [45]. As a new extension to the LDA model, the ratings were leveraged to provide valued side information and to further lead the model regarding inference on the singular pairs of sentiment and topics [44].

The categories of rating roughly corresponded to related labels in the dataset by López et al. [47] for this study, the category-to-topic mapping that was created included the technical (knowledgeability), systems (staff) and interpersonal (helpfulness).

4.4.2 Experiment and Analysis

The model for this study utilized two extensions to the LDA. The baseline model minus extensions was denoted by B, the model without word priors from labelled data was denoted by W, the model with document priors from ratings from users was denoted by R, and the full model with both extensions was denoted by WR [44].

The analysis also involved comparing the numbers of distributions of words against the LDA. For the iterations, the gradient ascent step was of size 10^{-3} and the Gaussian prior over the parameters was obtained. The LDA ran for similar number of repetitions to allow for the optimization of Dirichlet hyper-parameters for likelihood [41].

In the evaluation, the results obtained were from the default 3-fold cross-validation with each fold allowing for up to 6 inference trials to be performed through randomly initialized sampling sequences on the training set. The inferred parameters—with the lowest perplexity on the set that was held out was then selected. For inference on the held-out set, all parameters were then fixed, except for the ones that were document–specific.

4.5 RESULTS OF TOPIC-MODELING

The corpus turned out to have 84,292 files which contains text-reviews for each specialty. The reviews were generally short, with an average of 20 words for every review. Approximately 35% of the reviews were found to be one-liners. The breakdown of the online review's dataset, listing the number of reviews per physician for each of the top 20 practices is given as in Table 3;

Table 3: A breakdown of the online review dataset, with the number of reviews for each of the top twenty specialties, overall reviews and the number of rated healthcare providers.

Specialty	**# of Reviews**	**# of rated healthcare providers**
Acupuncturist	5634	773
Family Doctor / G.P.	139084	13614
Dentist	74342	7802
Chiropractor	27089	3268
Plastic / Cosmetic Surgeon, Physician	14287	414
Gynecologist (OBGYN)	12444	1007
General Surgeon	12076	1160
Optometrist	9915	1431
Orthopedic Surgeon	8884	773
Pediatrician	8740	918
Orthodontist	6905	345
Psychiatrist	6823	1137
Cardiologist	6422	934
Ophthalmologist	5991	636
Oncologist / Hematologist	5071	790
Naturopath	4814	801
Emergency Room Doctor	4767	799
Internist / Geriatrician	4316	769
Podiatrist	4185	298
Urologist	4146	335
Psychologist	4062	969
Total	3,69,997	38,973

Table 4 shows the chosen specialties for which LDA model has been run.

Table 4: A sample of the online review dataset, with the number of reviews for each of the specialties.

Specialty	# of Reviews
Acupuncturist	5634
Cardiologist	17389
Dentist	46987
Family Doctor / G.P.	139084
Chiropractor	27089
Plastic / Cosmetic Surgeon, Physician	14287
Gynecologist (OBGYN)	12444
General Surgeon	12076
Optometrist	9915
Orthopedic Surgeon	8884
Pediatrician	8740
Orthodontist	6905
Psychiatrist	6823
Ophthalmologist	5991
Oncologist / Hematologist	5071
Naturopath	4814
Emergency Room Doctor	4767
Internist / Geriatrician	4316
Podiatrist	4185
Urologist	4146
Psychologist	4062

Table 5: LDA words by Specialties

Specialty	Category	LDA Words
Acupuncturist	**Skills**	Treatment, Help, Accupunture
	Pain	Pain, Tell, Feel
	Knowledge	Good, Tell, Knowledgeable, Caring
Cardiologist	**Manners**	Doctor, Good, Care
	Friendly	Thank, Highly, Life
	Attention	Time, Explain, Question, Answer
Dentist	**Anecdotal**	Staff, Great
	Service	Concern, Long, Time
	Recommendation	Friendly, Care, Good, Recommend
Family Doctor / G.P.	**Compassion**	Doctor, Good, Care
	Schedule	Appointment, Time, Patient
	Friendly	Elle, Decin, Tra
Chiropractor	**Staff**	Great, Staff, Dentist
	Friendly	Tra, Elle, Decin
	Helpfulness	Doctor, Helpful, Care
Plastic / Cosmetic Surgeon, Physician	**Procedure**	Surgery, Result, Staff
	Schedule	Doctor, Time, Nose
	Positive	Surgery, Look, Year
Gynecologist (OBGYN)	**Interval**	Time, Doctor, Wait
	Advise	Tell, Surgery, Say
	Perceive	Doctor, Feel, Time
General Surgeon	**Expert**	Surgery, Doctor, Time
	Care	Good, Care, Great

	Disposition	Pour, Tra, Une
Specialty	**Category**	**LDA Words**
Optometrist	**Recommend**	Recommend, Staff, Eye
	Professionalism	Great, Time
	Compassion	Personable, Appointment, Exam
Orthopedic Surgeon	**Immense**	Surgery, Great, Knee
	Torment	Pain, Tell
	Schedule	Doctor, Time, Patient
Pediatrician	**Considerable**	Recommend, Foot, Staff
	Staff	Foot, Pain, Surgery
	Friendly	Time, Tell, Say
Orthodontist	**Helpfulness**	Tooth, Brace, Time
	Procedure	Staff, Great, Recommend
	Schedule	Make, Patient, Feel
Psychiatrist	**Positive**	Tell, Time, Medication
	Staff	Doctor, Good, Help
	Helpfulness	Help, Feel, Talk
Ophthalmologist	**Schedule**	Wait, Time, Doctor
	Procedure	Eye, Surgery, Year
	Staff	Staff, Doctor, Knowledgeable
Oncologist / Hematologist	**Compassion**	Doctor, Time, Care
	Skills	Treatment, Doctor, Patient
	Schedule	Time, Ask, Make
Naturopath	**Recommend**	Recommend, Health, Care
	Schedule	Time, Doctor, Make
	Helpfulness	Help, Year, Issue

Specialty	Category	LDA Words
Emergency Room Doctor	**Professionalism**	Doctor, Good, Time
	Knowledgeable	Knowledgeable, Feel, Family
	Positive	Elle, Pour, Decin
Internist / Geriatrician	**Staff**	Doctor, Say, Tell
	Helpfulness	Doctor, Care, Good
	Schedule	Time, Question, Recommend
Podiatrist	**Recommend**	Recommend, Foot, Staff
	Procedure	Foot, Pain, Surgery
	Schedule	Time, Tell, Say
Urologist	**Doctor**	Doctor, Question, Explain
	Schedule	Doctor, Time, Office
	Appointment	Appointment, Good, Make
Psychologist	**Helpfulness**	Help, Recommend, Feel
	Prescribe	Child, Say, Report
	Understanding	Experience, Son, Doctor

While assessing the inferred aspects for the various specialties, many of the aspects were seen to be shared between the various specialties even though the details varied between them [46]. Some of the aspects were specific to only some of the specialties or were not strongly exhibited in others.

The LDA that was later chosen had up to 60 topics picked to get results after consideration of semantic consistency, the percentage of topics having perceptible meaning and exclusivity scores [42]. For the model, topics were considered meaningful once seven distinctive and most common words from one topic were associated to the topic of services from general practitioners. The model with 60 topics offered detailed insights into the general practitioner service experience. At the same time, it avoided the generation of numerous meaningless topics.

The topics that were generated with the chosen model were labeled according to the most projecting words in the topic [42]. Features obtained from reviewing texts with the model were found to relate to various healthcare consumer experiences. The topics also had various prevalence through the review of dataset from 5% to 1% of tokens in the dataset [46]. The details of features extracted from text reviews are as shown in the table below.

The LDA model that was used also allowed for topics to be compared in terms of similarity to one another. Two topics were considered to be similar once the choice of vocabulary that they represent was seen to be the same [42]. Consequently, the topics were different if they had a few common words present [46]. The relative similarity between topics was done through linking the topics in the table above on a two-dimensional plane, with the distance apart being computed with cosine similarity.

In addition, the recurrence of terms like staff, help, knowledge, skill, in the reviews were related to how the reviewers rated their doctor's service experience in the numeric responses of a Likert scale.

The LDA further discovered several interesting insights. The top two most negative tokens from the model analysis turned out to be "arrogant" and "annoying." They highlighted the importance of communication and interpersonal skills. *Poor communication skills* turned out to be one of the most negative aspects of healthcare consumer care. Positive remarks, on the other hand, were dominated by superlatives such as "respectful" and "competent." The words that were associated with the topics generally matched the expectations, with interpersonal skills topics, including words like "caring" and "manners [41, 42, 43]." Technical topics contained words associated with "surgery" and other operations. System topics, on the other hand, contained words about the hospital and the doctors office, with words such as "staff", "appointments," and "booking."

The tables 5 to 8 represent the specialties. The important aspects inferred for each of the datasets broken down by specialties. For each aspect, the label is in bold which I have manually determined. The underlined words are the most frequent words determined by LDA for that aspect. The (possible) sample sentence in italics are extracted from the reviews and contains words associated with the aspect.

The characteristic of a good topic model is that it will have non-overlapping, large blobs representing each topic. This seems to be the case for the model output.

Table 9 presents a breakdown of the province, number of doctors, and the number of text-reviews that have been analyzed using the LDA model. Also, please note that to maintain consistency and avoid issues like ip blocking and losing internet connection to the cloud platform, the number of doctors for Ontario (ON) and British Columbia (BC) have been reduced to run the model successfully.

Table 6: A breakdown of the province, with the number of doctors, and number of text-reviews.

Province	Number of Doctors	Number of Text-Reviews
Alberta (AB)	9,165	183,342
Ontario (ON)	9,701	301,037
British Columbia (BC)	8,889	137,342
Manitoba (MB)	3,191	61,171
Nova Scotia (NS)	396	43,594
New Brunswick (NB)	1,661	28,195
Prince Edward Island (PE)	295	4,901
NewFoundland&Labrador (NL)	989	14,525
Saskachewan (SASK)	3,620	51,244
Quebec (QC)	12,830	167,467

The following tables represent the geographic location i.e. Canadian Provinces, including all the specialties. The important aspects inferred for each of the datasets broken down by geographic location. For each aspect, the label is in bold which was manually determined. The (possible) sample sentence, in italics, are extracted from the reviews and contains words associated with the aspect.

Table 7: LDA Words by Provinces

Provinces	Category	LDA Words
Alberta (AB)	**Excitement**	Doctor, Good, great
	Satisfaction	Tell, Surgery, Pain
	Friendly	Good, Helpful
Ontario (ON)	**Appointment**	Time, Wait, Appointment
	Recommendation	Doctor, Patient, Good
	Professionalism	Surgery, Pain, Surgeon
British Columbia (BC)	**Amazing**	Doctor, Patient, Time
	Service	Good, Care, Great
	Thorough	Filling, Use, Professional
Manitoba (MB)	**Great**	Doctor, Good, Care, Great
	Patient	Test, Patient, Tell
	Professionalism	Ask, Say, Tell
Nova Scotia (NS)	**Professionalism**	Amazing, Good, Great
	Service	Nice, Concern, Long, Appointment
	Recommendation	Family, Healthcare Consumer, Feel
New Brunswick (NB)	**Attitude**	Tell, Doctor
	Doctor	Doctor, Good, Care
	Appointment	Time, Make, Feel
Prince Edward Island (PEI)	**Recommendation**	Tell, Say
	Care	Pain, Doctor, Care
	Staff	Great, Staff, Excitement
NewFoundland&Labrador (NL)	**Recommendation**	Doctor, Tell, Say
	Schedulling	Time, Make, Appointment
	Humane	Good, Great, Care
Saskachewan (SASK)	**Knowledgeable**	Tell, Make, Doctor
	Recommendation	Recommend, Professional
	Wit	Work, Understand

Quebec (QC)	**Que**	Pas, Pour, Que
	Experience	Old, Condition, Experience
	Recommendation	Doctor, Good, Time

A breakdown of the territories, with the number of doctors, and number of text-reviews.

Table 8: LDA Words by Territories

Territories	Number of Doctors	Number of Text-Reviews
Northwest Territories (NT)	**Treatment**	Doctor, Treat
	Recommendation	Tell, Say
	Professionalism	Doctor, Good, Care
Yukon	**Knowledgeable**	Doctor, Help, Care
	Emergency	Emergency, Appear
	Nonprofessional	Patient, Doctor
Nunavut (NU)	**Professionalism**	Doctor, Care
	Recommendation	Time, Good
	Staff	Staff, Work

The LDA further discovered several interesting salient patterns. The top two most negative tokens from the model analysis turned out to be "arrogant" and "annoying." They highlighted the importance of communication and interpersonal skills. Poor communication skills turned out to be one of the most complained about aspects of healthcare consumer care. Positive remarks, on the other hand, were dominated by superlatives such as "respectful" and "competent**".** The words that were associated with the topics generally matched the expectations, with interpersonal topics, including words like "caring" and "manners [2, 3, 4]." Technical topics contained words associated with surgery and other operations. System topics, on the other hand, contained words about the hospital and the doctors' office, with the words being such as "staff," "appointments," and "booking."

Appendix D contains further visualizations for number of topics, obtained through pyLDAvis, both for Specialties and Geographic Locations (Provinces/Territories) respectively. Google translator [48] is used to translate French to English in order to place the review comments under the correct categorization.

From the results, it turns out that implementing such bottom-up approach proved to be promising, especially in identifying the common aspects of providers that health consumers often review as well as those that are specific to each geo-location and specialty.

CHAPTER 5 DISCUSSION

5.1 ONLINE DOCTORS REVIEW

With the data at the national level over the span for the period from September 2013 to January 2019, using the physician-rating website, a total of 92,296 unique rated healthcare providers, 10,94,180 ratings for geography (provinces and territories) and 13,36,039 ratings for specialties were found to be rated online. The overall ratings were positive, above 3, which conforms to physician-rating results from previous studies. Based on the results presented in chapter 4, Section 4.1, it turned out that there were differences in the ratings in terms of specialty as well as geographical locations—which was demarcated by provinces. In line with previous studies, this research also described the landscape of doctors' ratings within the country.

The results showed that healthcare providers in some of the specialties, such as plastic surgery, family doctors, and dentistry, had a higher likelihood of being rated than others such as podiatrists, radiation oncologists, and osteopath. The study also realized that various factors touching on either healthcare providers or healthcare consumers were likely to contribute to such differences in ratings as was exhibited [50]. Some of the differences that were found included healthcare consumer population and healthcare consumer expectations. For instance, while recognition of recovery from addiction might be rateable for psychiatrists, recovery from physical injury is not. Elsewhere, while surviving radiology might be rateable for radiation oncologists, recognition of milestones towards emotional recovery might not be straightforwardly rateable. Additionally, there was a high likelihood that complex interactions between healthcare consumers' preconceived perception of the healthcare providers, especially regarding the healthcare providers' performance as well as their resulting satisfaction. The interaction was driven by the expectation-disconfirmation theory [50].

The results from this research add more information to work done by previous researchers, with quality of rating being shown to be similar for healthcare providers in surgical specialties, obstetrics, gynecology as well as primary care. The ratings, however, turned out to be significantly different for a category of other healthcare providers in radiology, pathology, and anesthesiology. The research, however, tended to differ with previous studies that showed generalists to have favorable ratings as the difference showed both in quantity and quality from the respective subspecialist groups.

Furthermore, the results of this research study can be used by the healthcare authorities and stakeholders several ways. For instance, the feedback about the staff of a GP (General Practitioner) could signal for more rebust training in customer services. The cultural aspects are intertwined with the healthcare system of a geography, i.e. availability of doctors in a proxemic zone according to the increase in population of that region.

Looking at the rating by geographical location (chapter 4, section 4.1), this research found minimal differences in the likelihood of positive rating as it seemed unlikely that the quality of healthcare providers' performance differed vastly with region. The minimal difference by province was most likely owed to the continuing education and the national accreditation standard. The study, however, realized that Atlantic provinces seemed to have a higher likelihood of physician rating than the West coast, upholding results from previously conducted research studies.

Although there was minimal difference in the mean overall rating by province, the small differences that were realized came due to various reasons such as the difference in accessibility due to location. *The study tended to agree with the hypothesis that the scarce healthcare providers get*, *the more appreciative the consumers become*. As a result, provinces that had scarce healthcare providers were more likely to have high mean overall rating than those where healthcare providers were in plenty. Therefore, limited accessibility tended to impose a bias on the rating in a favorable manner.

Furthermore, factors that influenced the rating of healthcare providers within the provinces were population density. Provinces that had relatively higher economic prosperity tended to have lower healthcare consumer satisfaction as compared to those that had relatively lower prosperity. The findings, were, however, contrary to previously held theories by scholars such as Grigoroudis, which claims that high economic prosperity translates to high healthcare consumer satisfaction. Additionally, other cultural or sociological phenomena across the regions also influenced the various consumer preference. The explanation for such phenomena is, however, yet to be realized as there is limited research regarding the variability of the rating of healthcare providers according to geographic practice locations.

5.2 LIMITATIONS

Following are several notable limitations in this research study.

(1) Data anomalies: I have observed many anomalies in the data scraped from RateMDs.com. For instance, while pre-processing the data for geographic analysis according to province, I came across irrelevant information. Talking about the Canadian provinces; Quebec has been called by QC, Queb+Ec, Quebec. Similarly, British Columbia has been referred by BC, Vancouver, Victoria, Surrey; Alberta has been referred by Alberta, AB; Ontario has been referred by ON, Ontario, Vaughan, Brampton.

(2) Access to computer and internet: Another notable limitation is that the online physician ratings could not be generalized, as it assumes healthcare consumers have access to the computer and internet. Those without access may not be able to rate.

(3) Duplicate profiles: As the healthcare consumers have the mandate to be allowed to see the healthcare provider's profile, there is no mechanism in place to control the duplicate profiles under different names.

(4) Anonymity of raters: This provide a loophole for none-authentic ratings.

(5) Influence on the future ratings: For instance, a user that logs in to post their rating may have their original inclination influenced by the previously posted average rating [5]. As a result, the rating results are bound to have a relatively steady trajectory regardless of the healthcare providers' performance.

(6) Computing resources: – The study proposes the use of high computing resources, such as CPU, to run the model in a timely manner.

(7) Fine tunning: The LDA model needs a lot of fine tunning, which means time and energy commitment.

(8) Topics can not be influenced: One of the limitations of the model is that the choice of the topics can not be influenced.

(9) Human interpretation: The LDA model needs human interpretation based on the probabilities assigned to the words.

(10) Weakness of the Data collection: The website allowed for registration as a doctor without an input validation mechanism. As a result, the dataset obtained contained

numerous incoherent names of provinces and cities, which made the data cleaning a tedious task, given that the number of records were extremely high. As a way forward, the website should have adequate input validations that will prevent unwanted entries. For instance, since the names of cities are already known, it would be proper that they are chosen from a list instead of being manually typed in.

CHAPTER 6 CONCLUSION

6.1 CONCLUSION

Feedback, in the form of reviews, are an important source of information for both the healthcare providers and consumers. Efforts should be expended to ensure the reliability of such reviews. Websites, such as RateMDs, are one such sources of reviews that allow healthcare consumers to rate their respective healthcare providers both on a numerical scale and using qualitative writing (comments). Previous research studies have shown significant growth of physician-rating websites. The results of this study showed that the physician ratings were generally positive for various specialties. However, there is still a need to further explore the significant difference in rating according to specialties, such as dentistry and surgery which have a higher likelihood of being rated as compared to other specialties such as podiatry.

Furthermore, it can be concluded from the second part of the research i.e. topic-modeling using LDA, that analyzing the text-review of the healthcare providers can offer valuable insights to healthcare stakeholders. We presented a complimentary approach of Unsupervised Machine Learning method to identify salient recurring aspects from a large corpus of the reviews authored by the healthcare consumers. This qualitative aspect of the information gleaned from the reviews could help healthcare authorities in making strategic decisions.

6.2 FUTURE WORK

While conducting this research study, the following questions emerged that might warrant further exploration and research.

(1) This research explored the text-reviews for doctors in Canada. It would be interesting to see a comparison between the reviews for doctors in USA and Canada.

(2) In this research, the texts corpus scraped from RateMDs.com were only the length of a paragraph i.e. short documents. It might be worth exploring if the result changes when scraping a full-length article.

(3) Another interesting research area will be to explore the model being applied to streaming data or perhaps evaluate a text corpus with altering the ‘Search Params’.

(4) Although, meta-data is considered as the data labels for Unsupervised Learning, the text-reviews scraped from RateMDs.com could be utilized in a Supervised Learning setup by manually tagging the data. In addition, the assumption of accuracy could also be verified.

(5) A future research avenue might be to investigate if the LDA model could be deployed and visualized using other Python libraries.

(6) Non-parametric topic models, such as Hierarchical Dirichlet Process (HDP), could also be applied to the text corpus.

(7) Further, it is yet to be explored which programming language, i.e. Python or R, is more suiteable for topic-modeling.

(8) The model's parameters, i.e. the number of topics, (k) and the prior parameters, (Alpha and Beta), needs to be explored to a greater extend for reliability and validity.

(9) Finally, future research could explore the feasibility of mapping the topics identified in a corpus to social media data, such as that found on facebook and twitter.

BIBLIOGRAPHY

[1] A. M. Holliday, A. Kachalia, G. S. Meyer, and T. D. Sequist, "Physician and Patient Views on Public Physician Rating Websites: A Cross-Sectional Study," *J. Gen. Intern. Med.*, vol. 32, no. 6, pp. 626–631, Jun. 2017.

[2] F. S. Bäumer, J. Kersting, V. Kuršelis, and M. Geierhos, "Rate Your Physician: Findings from a Lithuanian Physician Rating Website," Springer, Cham, 2018, pp. 43–58.

[3] "Access Data and Reports | CIHI." [Online]. Available: https://www.cihi.ca/en/access-data-and-reports. [Accessed: 16-Oct-2019].

[4] "Terms of Use - RateMDs." [Online]. Available: https://www.ratemds.com/about/terms/. [Accessed: 08-Aug-2019].

[5] J. J. Liu, J. J. Matelski, and C. M. Bell, "Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study," *J. Med. Internet Res.*, vol. 20, no. 3, p. e76, Mar. 2018.

[6] M. Emmert, L. Sauter, L. Jablonski, U. Sander, and F. Taheri-Zadeh, "Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period," *J. Med. Internet Res.*, vol. 19, no. 7, pp. 1–13, 2017.

[7] T. Lagu, N. S. Hannon, M. B. Rothberg, and P. K. Lindenauer, "Patients' evaluations of health care providers in the era of social networking: An analysis of physician-rating websites," *J. Gen. Intern. Med.*, vol. 25, no. 9, pp. 942–946, 2010.

[8] B. Kadry, L. F. Chu, B. Kadry, D. Gammas, and A. MacArio, "Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating," *J. Med. Internet Res.*, vol. 13, no. 4, 2011.

[9] G. G. Gao, J. S. McCullough, R. Agarwal, and A. K. Jha, "A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period.," *J. Med. Internet Res.*, vol. 14, no. 1, pp. 1–11, 2012.

[10] E. Levine and E. Domany, "Resampling Method for Unsupervised Estimation of Cluster Validity," *Neural Comput.*, vol. 13, no. 11, pp. 2573–2593, Nov. 2001.

[11] T. Lagu *et al.*, "Website characteristics and physician reviews on commercial physician-rating websites," *JAMA*, vol. 317, no. 7, p. 766, 2017.

[12] S. Brody and N. Elhadad, "Human An Unsupervised Aspect-Sentiment Model for Online Reviews."

[13] S. Brody and N. Elhadad, "Detecting salient aspects in online reviews of health providers.," *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2010, pp. 202–6, Nov. 2010.

[14] I. Titov and R. Mcdonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization."

[15] M. Emmert, F. Meier, A. K. Heider, C. Dürr, and U. Sander, "What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website," *Health Policy (New. York).*, vol. 118, no. 1, pp. 66–73, 2014.

[16] M. Emmert and F. Meier, "An analysis of online evaluations on a physician rating

website: Evidence from a german public reporting instrument," *J. Med. Internet Res.*, vol. 15, no. 8, 2013.

[17] W. S. Chou, Y. M. Hunt, E. B. Beckjord, R. P. Moser, and B. W. Hesse, "Social media use in the United States: implications for health communication.," *J. Med. Internet Res.*, vol. 11, no. 4, p. e48, Nov. 2009.

[18] H. R. Rubin, B. Gandek, W. H. Rogers, M. Kosinski, C. A. McHorney, and J. E. Ware, "Patients' ratings of outpatient visits in different practice settings. Results from the Medical Outcomes Study.," *JAMA*, vol. 270, no. 7, pp. 835–40, Aug. 1993.

[19] J. Segal, M. Sacopulos, V. Sheets, I. Thurston, K. Brooks, and R. Puccia, "Online Doctor Reviews: Do They Track Surgeon Volume, a Proxy for Quality of Care?," *J. Med. Internet Res.*, vol. 14, no. 2, p. e50, Apr. 2012.

[20] D. A. Hanauer, K. Zheng, D. C. Singer, A. Gebremariam, and M. M. Davis, "Parental awareness and use of online physician rating sites.," *Pediatrics*, vol. 134, no. 4, pp. e966-75, Oct. 2014.

[21] C. Frost and A. Mesfin, "Online Reviews of Orthopedic Surgeons: An Emerging Trend," *Orthopedics*, vol. 38, no. 4, pp. e257–e262, Apr. 2015.

[22] E. Grigoroudis, G. Nikolopoulou, and C. Zopounidis, "Customer satisfaction barometers and economic development: An explorative ordinal regression analysis," *Total Qual. Manag. Bus. Excell.*, vol. 19, no. 5, pp. 441–460, May 2008.

[23] M. M. Galizzi *et al.*, "Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London.," *BMJ Open*, vol. 2, no. 6, p. e001493, Jan. 2012.

[24] Paul H. Keckley, *2011 Survey of Health Care Consumers in the United States: Key Findings ... - Paul H. Keckley - Google Books*. .

[25] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects.," *Science*, vol. 349, no. 6245, pp. 255–60, Jul. 2015.

[26] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jan. 2008.

[27] G. Kennedy *et al.*, "Technology Solutions to Combat Online Harassment," in *Proceedings of the First Workshop on Abusive Language Online*, 2017, pp. 73–77.

[28] B. Gambäck and U. Kumar Sikdar, "Using Convolutional Neural Networks to Classify Hate-Speech."

[29] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A Convolutional Neural Network for Modelling Sentences," Association for Computational Linguistics.

[30] G. Ganu, Y. Kakodkar, and A. Marian, "Improving the quality of predictions using textual information in online user reviews," *Inf. Syst.*, vol. 38, no. 1, pp. 1–15, Mar. 2013.

[31] X. Zhang and Y. Lecun, "Text Understanding from Scratch."

[32] Y. Kim, "Convolutional Neural Networks for Sentence Classification."

[33] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach," *Cognit. Comput.*, vol. 7, no. 4, pp. 487–499, Aug. 2015.

[34] H. Liu and P. Singh, "ConceptNet-a practical commonsense reasoning tool-kit," 2004.

[35] A. Wawer, "Towards Domain-Independent Opinion Target Extraction," in *2015*

IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1326–1331.

[36] Nikki Castle, “Supervised vs. Unsupervised Machine Learning.” [Online]. Available: https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms. [Accessed: 07-Aug-2019].

[37] E. Vargiu and M. Urru, “Exploiting web scraping in a collaborative filtering-based approach to web advertising,” *Artif. Intell. Res.*, vol. 2, no. 1, 2013.

[38] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, “Web scraping technologies in an API world,” *Brief. Bioinform.*, vol. 15, no. 5, pp. 788–797, Sep. 2014.

[39] L. Sobin and P. Goyal, “Trends of Online Ratings of Otolaryngologists,” *JAMA Otolaryngol. Neck Surg.*, vol. 140, no. 7, p. 635, Jul. 2014.

[40] D. M. Blei, A. Y. Ng, and J. B. Edu, “Latent Dirichlet Allocation Michael I. Jordan,” 2003.

[41] “Dr. Michael Robbins - Markham, ON - Chiropractor Reviews & Ratings - RateMDs.” [Online]. Available: https://www.ratemds.com/doctor-ratings/3540782/Dr-Michael-Robbins-Markham-ON.html. [Accessed: 07-Sep-2019].

[42] J. G. Merrell, B. H. Levy, and D. A. Johnson, “Patient Assessments and Online Ratings of Quality Care: A ‘Wake-Up Call’ for Providers,” *Am. J. Gastroenterol.*, vol. 108, no. 11, pp. 1676–1685, Nov. 2013.

[43] F. Hogenboom, F. Frasincar, U. Kaymak, F. de Jong, and E. Caron, “A Survey of event extraction methods from text for decision support systems,” *Decis. Support Syst.*, vol. 85, pp. 12–22, May 2016.

[44] R. Kowalski, “Patients’ written reviews as a resource for public healthcare management in England,” *Procedia Comput. Sci.*, vol. 113, pp. 545–550, Jan. 2017.

[45] Z.-Y. Niu, D. Ji, and C. L. Tan, “I2R: Three Systems for Word Sense Discrimination, Chinese Word Sense Disambiguation, and English Word Sense Disambiguation.” pp. 177–182, 2007.

[46] X. Wei and W. B. Croft, “LDA-Based Document Models for Ad-hoc Retrieval,” 2006.

[47] A. López, A. Detz, N. Ratanawongsa, and U. Sarkar, “What Patients Say About Their Doctors Online: A Qualitative Content Analysis,” *J. Gen. Intern. Med.*, vol. 27, no. 6, pp. 685–692, Jun. 2012.

[48] “Google Translate.” [Online]. Available: https://translate.google.com/?rlz=1C1GGRV_enCA757CA757&um=1&ie=UTF-8&hl=en&client=tw-ob#fr/en/Le meilleur mÃƒÂ©decin de famille que lon peut avoir! [Accessed: 16-Oct-2019].

APPENDIX A Tools Used

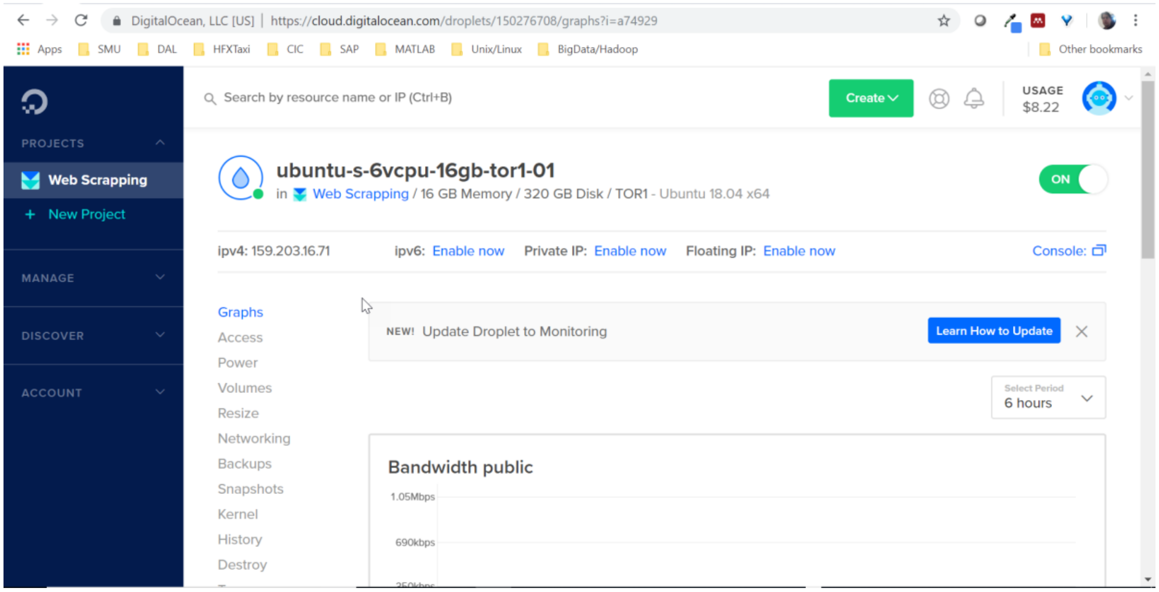

Figure 7: The Digital Ocean (Cloud) platform

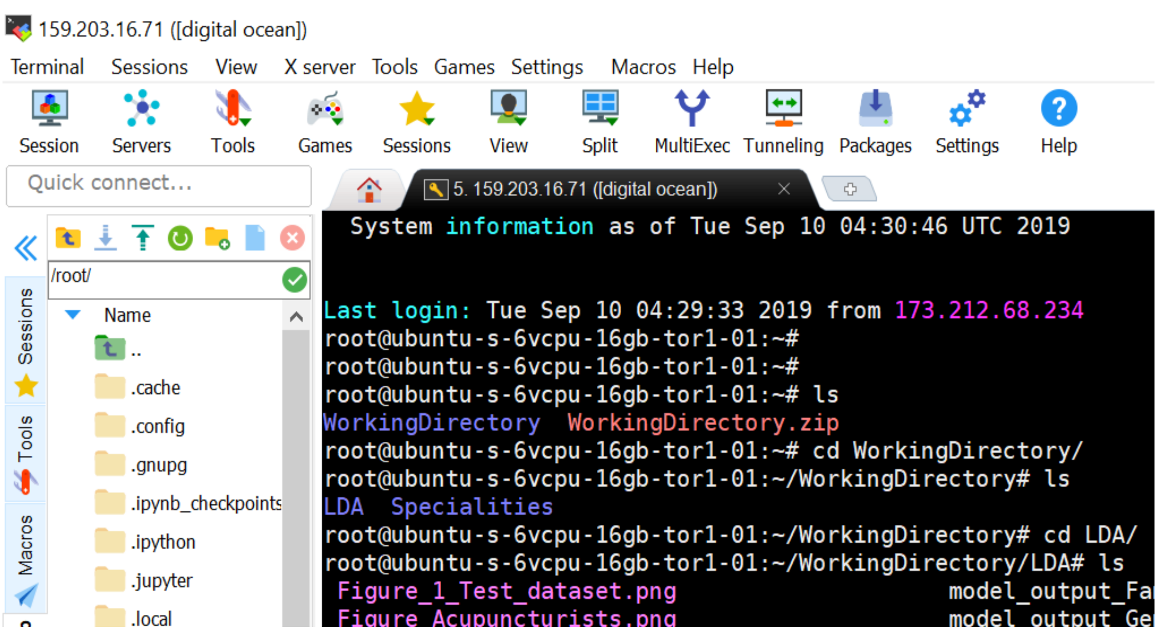

Figure 8: Connecting to Digital Ocean via MobaXterm

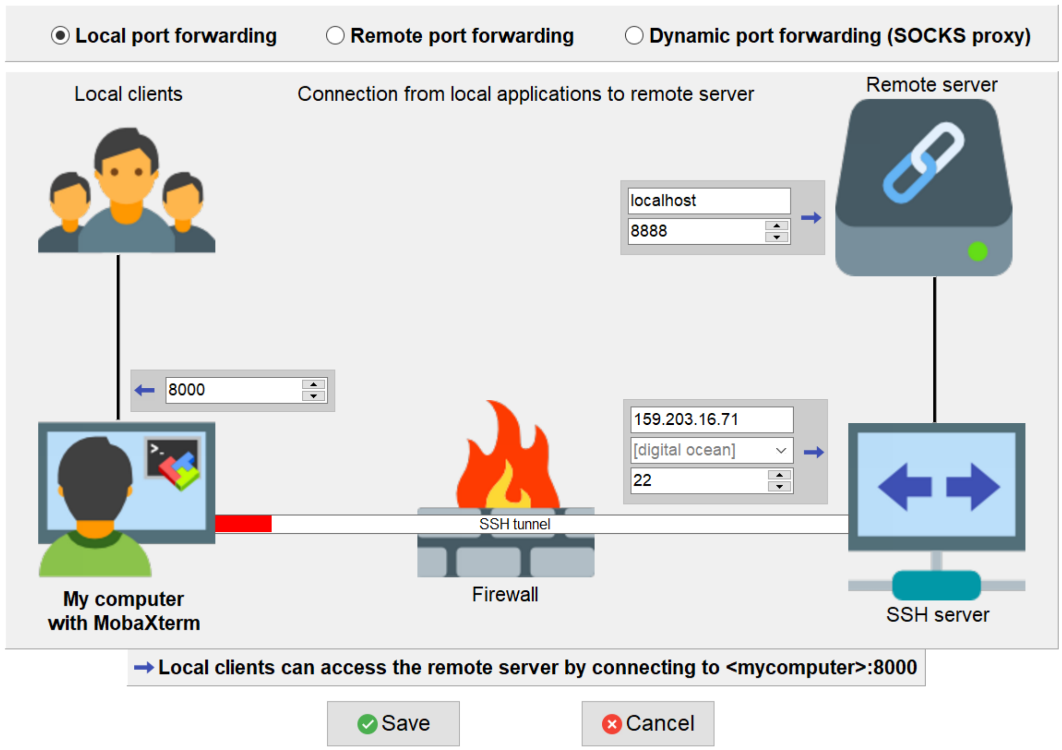

Figure 9: Tunneling to Jupyter Notebook on Digital Ocean

Figure 10: RateMDs.com/best-doctors/?country=ca

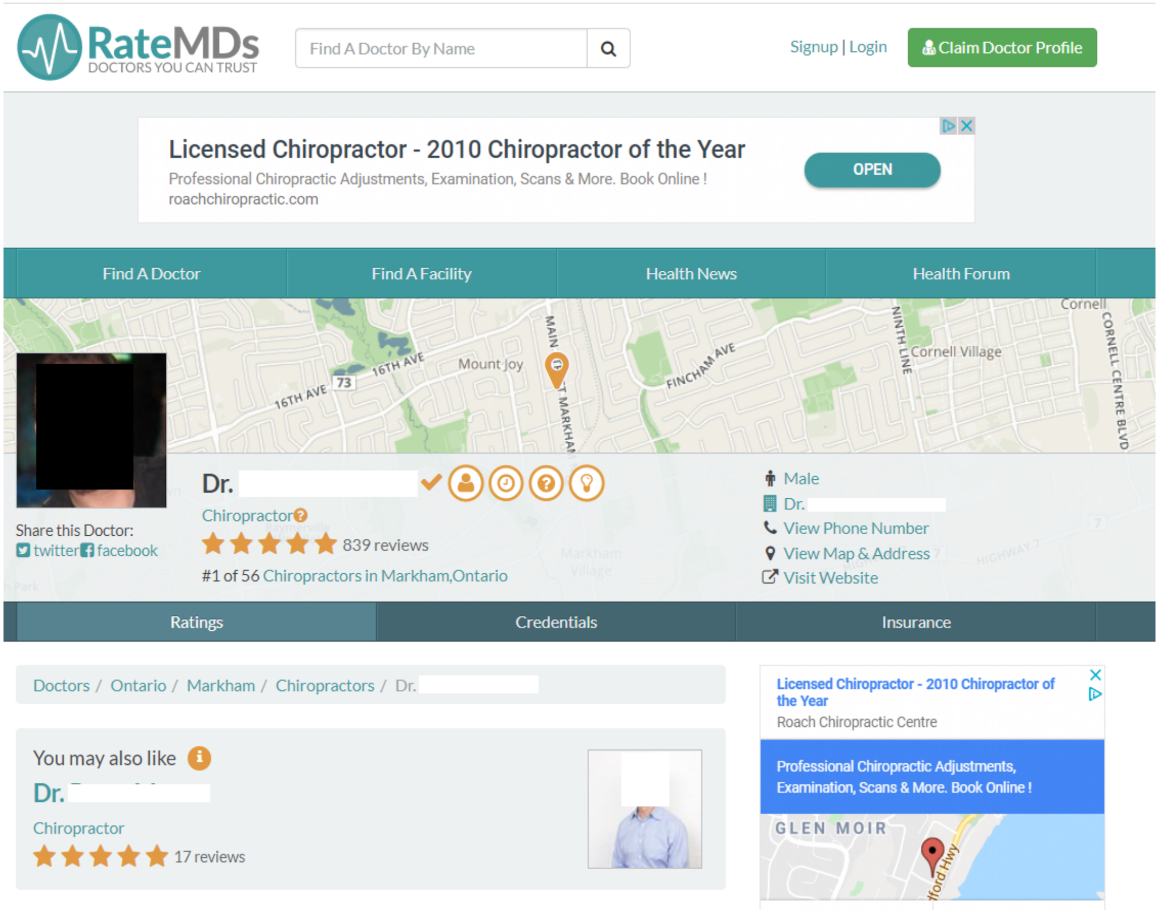

Figure 11: Doctor's Personal Webpage

APPENDIX B Complete program execution

(1). This section contains python **code for 'Acupuncturists' specialty:**

```
from sklearn.externals import joblib
import datetime
from reading_data import get_read_specialization_data

# Clusting data
# Run in terminal or command prompt
# python3 -m spacy download en

import numpy as np
import pandas as pd
import re
import nltk
import en_core_web_sm
import spacy
import gensim

# Sklearn
from sklearn.decomposition import LatentDirichletAllocation,
TruncatedSVD
from sklearn.feature_extraction.text import CountVectorizer,
TfidfVectorizer
from sklearn.model_selection import GridSearchCV
import pickle
from tqdm import tqdm


# Plotting tools
import pyLDAvis
import pyLDAvis.sklearn
import matplotlib.pyplot as plt


path = r"../Specialities/reviews"
specialization = "Acupuncturists"

data = get_read_specialization_data(path, specialization)

data.columns=["reviews"]
print(len(data) )
data['reviews'].replace('\n', np.nan, inplace=True)
data.dropna(subset=['reviews'], inplace=True)
print(data.head(100))
len(data)

100%|██████████| 773/773 [00:00<00:00, 26153.71it/s]
```

```
Reading the review from DRs files
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7390 entries, 0 to 7389
Columns: 1 entries, 0 to 0
dtypes: object(1)
memory usage: 57.8+ KB
None
7390
                                                reviews
0    He is best acupuncturist. Smart and loyal to p...
1    Dr. xianqi Wu? Wow! He is the best in town! I ...
2    He is unbelievably smart and talented. I didn'...
4    Also the atmosphere is very relaxing and quite...
5    Very good Acupuncturist and herbal medicine pr...
6    When my husband and I were trying to get pregn...
7    Gloria is an extremely talented and empathetic...
8    Gloria is very helpful and is kind and very pr...
9    I cannot say enough good things about Gloria a...
10   Gloria Chu is very professional and knowledgea...
11   I had a severe car accident in May of 2010. My...
12   Cedric Kam Tat Cheung is a professor and acupu...
13   I have been a healthcare consumer of Dr. Lee for nerve pai...
14   I have been dealing with chronic headache and ...
15   I began seeing Dr. Lee just a couple of months...
16   I have been an healthcare consumer of Jay Lee for 4 years....
18                                           Lillie C.\n
19   I was having severe Migraine headaches and wit...
20   I have had acupuncture treatments in the past ...
21   Dr. Jae Seok Lee is treating me for the sympto...
22   I've seen few different acupuncturists to get ...
23   I seen Dr. Lee after having a young female chi...
25   I found Dr. Lee to be incredibly knowledgeable...
27   It was my first experience with acupuncture, h...
29   I am so grateful that it was Dr. Lee that I fo...
31   Anyways, the bottom line is......after living ...
33   All I know is that Dr. Lee is extremely good a...
34   I have gone to Lana for many years after tryin...
35   She was good at figuring out what I had and wh...
36   This is a very good acupunturist, very effecti...
..                                                 ...
88   2. I have been treated in every treatment room...
90   3. The reception staff is lovely and approacha...
92   It would seem to me that while this healthcare consumer ma...
93   When I arranged to see Xiaolan, someone else c...
94   Amazing!!! I've had a jaw problem for the last...
95   Some people learn from books, take what they h...
97   If you ask her questions, she answers them and...
99   She is amazing at what she does and very compa...
102  One of the reviews below says Dr. Zhao takes y...
103  My theory is that xioalan has limited availabi...
```

```
104  My husband and I and quite a few friends have ...
106  Dr Xiaolan actually has two degrees. She studi...
108  Before you go, you should know that sheâ€™s ve...
110  Hereâ€™s what happens: you will take off your ...
112  Sheâ€™ll ask you to take off your clothes, lea...
114  Dr Xiaolan may turn a heat lamp on, leave the ...
116  Afterwards you will probably be given herbs to...
118  By the way, a little note to my fellow reviewe...
119  I have been going to Xiaolan for almost 20 yea...
121  I have absolutely been able to manage my asthm...
123  They don't try to get you to do or take things...
125  They see me at my appointed time, on the dot, ...
126  I have been to the Xiaolan Health Centre MANY ...
127  At my first appointment, I was shocked when my...
128  After reading the reviews below, both positive...
129  Over 6 weeks, Xiaolan and her superb staff, fa...
130  I can't believe the degree of negative comment...
131  I originally booked my appointment with Xiaola...
132  I find it interesting that the majority of dis...
133  Xiolan Health Clinic feels cold and caring. I ...

[100 rows x 1 columns]


5635
```

```python
print("Print Data")
data.head(15)
text_list = data["reviews"].values.tolist()
data = text_list


# Cleaning Data
# Remove Emails
data = [re.sub('\S*@\S*\s?', '', sent) for sent in data]

# Remove new line characters
data = [re.sub('\s+', ' ', sent) for sent in data]

# Remove distracting single quotes
data = [re.sub("\'", "", sent) for sent in data]
print(data[:1])
```

```
<>:9: DeprecationWarning: invalid escape sequence \S
<>:12: DeprecationWarning: invalid escape sequence \s
<>:9: DeprecationWarning: invalid escape sequence \S
<>:12: DeprecationWarning: invalid escape sequence \s
<>:9: DeprecationWarning: invalid escape sequence \S
<>:12: DeprecationWarning: invalid escape sequence \s
```

```
<ipython-input-10-3527fe854bea>:9: DeprecationWarning: invalid escape
sequence \S
  data = [re.sub('\S*@\S*\s?', '', sent) for sent in data]
<ipython-input-10-3527fe854bea>:12: DeprecationWarning: invalid escape
sequence \s
  data = [re.sub('\s+', ' ', sent) for sent in data]


Print Data
['He is best acupuncturist. Smart and loyal to profession. ']
```

```
def sent_to_words(sentences):
 for sentence in sentences:
 # deacc=True removes punctuations
 yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))


data_words = list(sent_to_words(data))

print(data_words[:1])
```

```
[['he', 'is', 'best', 'acupuncturist', 'smart', 'and', 'loyal', 'to',
'profession']]
```

```
def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB',
'ADV']):
    """https://spacy.io/api/annotation"""
    texts_out = []
    for sent in tqdm(texts):
        doc = nlp(" ".join(sent))
        texts_out.append(" ".join([token.lemma_ if token.lemma_ not in
[
 '-PRON-'] else '' for token in doc if token.pos_ in allowed_postags]))
    return texts_out


# Initialize spacy 'en' model, keeping only tagger component (for
efficiency)
# Run in terminal: python3 -m spacy download en
nlp = en_core_web_sm.load()


# Do lemmatization keeping only Noun, Adj, Verb, Adverb
data_lemmatized = lemmatization(data_words, allowed_postags=[
                                'NOUN', 'ADJ', 'VERB', 'ADV'])

print(data_lemmatized[:2])
```

```
100%|██████████| 5635/5635 [01:14<00:00, 75.81it/s]

['be good acupuncturist smart loyal profession', 'xianqi wu be good
```

```
town dona even know where begin great thing have say family use clinic
health issue be be healthcare consumer listen sincerely care whole
health be highly skilled professional extremely compassionate address
concern answer question work entire process treat respect dignity have
only positive feedback regard entire experience everything scheduling
treatment would text call follow be satisfied treatment receive would
highly recommend anyone need traditional chinese medicine therapy be
woman experience health issue such weight management high cycle
hormonal other issue be right be amazing service fee be affordable
outcome be priceless']
```

```
print("Start CountVectorizer")

vectorizer = CountVectorizer(analyzer='word',
                             min_df=10,                        #
minimum reqd occurences of a word
                             stop_words='english',             # remove
stop words
                             lowercase=True,                   #
convert all words to lowercase
                             token_pattern='[a-zA-Z0-9]{3,}',  # num
chars > 3
                             max_features=50000,               # max
number of uniq words
                             )
start = datetime.datetime.now()
print("Time of start", start)
data_vectorized = vectorizer.fit_transform(data_lemmatized)
end = datetime.datetime.now()
print("Time of start", end)
print("taken time", end-start)
```

```
Start CountVectorizer
Time of start 2019-07-08 18:15:41.589030
Time of start 2019-07-08 18:15:41.762622
taken time 0:00:00.173592
```

```
# Define Search Param
n_topics = [5, 7, 9]
search_params = {'n_components': n_topics, "learning_method": [
    "online"], 'learning_decay': [.5, .7, .9]}

# Init the Model
lda = LatentDirichletAllocation(n_jobs=-1, batch_size=128)

# Init Grid Search Class
model = GridSearchCV(lda, n_jobs=-1, param_grid=search_params,
verbose=5)

# Do the Grid Search
model.fit(data_vectorized)
```

```
Fitting 3 folds for each of 12 candidates, totalling 36 fits
```

```
GridSearchCV(cv='warn', error_score='raise-deprecating',
  estimator=LatentDirichletAllocation(batch_size=128,
  doc_topic_prior=None,
  evaluate_every=-1,
  learning_decay=0.7,
  learning_method='batch',
  learning_offset=10.0,
  max_doc_update_iter=100,
  max_iter=10,
  mean_change_tol=0.001,
  n_components=10, n_jobs=-1,
  perp_tol=0.1,
  random_state=None,
  topic_word_prior=None,
  total_samples=1000000.0,
  verbose=0),
  iid='warn', n_jobs=-1,
  param_grid={'learning_decay': [0.5, 0.7, 0.9],
    'learning_method': ['online'],
    'n_components': [5, 6, 7, 8]},
     pre_dispatch='2*n_jobs', refit=True, return_train_score=False,
     scoring=None, verbose=5)
```

```python
# Best Model
best_lda_model = model.best_estimator_

# Model Parameters
print("Best Model's Params: ", model.best_params_)

# Log Likelihood Score
print("Best Log Likelihood Score: ", model.best_score_)

# Perplexity
print("Model Perplexity: ", best_lda_model.perplexity(data_vectorized))

# Get Log Likelyhoods from Grid Search Output
log_likelyhoods_5 = []
log_likelyhoods_7 = []
log_likelyhoods_9 = []

cv_result = pd.DataFrame(model.cv_results_)
for index, gscore in cv_result.iterrows():
    if gscore["params"]["learning_decay"] == 0.5:
        log_likelyhoods_5.append(round(gscore["mean_test_score"]))
    if gscore["params"]["learning_decay"] == 0.7:
        log_likelyhoods_7.append(round(gscore["mean_test_score"]))
    if gscore["params"]["learning_decay"] == 0.9:
```

```
        log_likelyhoods_9.append(round(gscore["mean_test_score"]))

# Show graph
plt.figure(figsize=(12, 8))
plt.plot(n_topics, log_likelyhoods_5, label='0.5')
plt.plot(n_topics, log_likelyhoods_7, label='0.7')
plt.plot(n_topics, log_likelyhoods_9, label='0.9')
plt.title("Choosing Optimal LDA Model")
plt.xlabel("Num Topics")
plt.ylabel("Log Likelyhood Scores")
plt.legend(title='Learning decay', loc='best')
plt.show()


# Create Document - Topic Matrix
lda_output = best_lda_model.transform(data_vectorized)

# column names
topicnames = ["Topic" + str(i) for i in
range(best_lda_model.n_components)]

# index names
docnames = ["Doc" + str(i) for i in range(len(data))]

# Make the pandas dataframe
df_document_topic = pd.DataFrame(
    np.round(lda_output, 2), columns=topicnames, index=docnames)

# Get dominant topic for each document
dominant_topic = np.argmax(df_document_topic.values, axis=1)
df_document_topic['dominant_topic'] = dominant_topic

# Styling


def color_green(val):
    color = 'green' if val > .1 else 'black'
    return 'color: {col}'.format(col=color)


def make_bold(val):
    weight = 700 if val > .1 else 400
    return 'font-weight: {weight}'.format(weight=weight)
```

```
Best Model's Params:  {'learning_decay': 0.7, 'learning_method':
'online', 'n_components': 5}
Best Log Likelihood Score:  -250116.83779137878
Model Perplexity:  669.7453855423448
```

Png

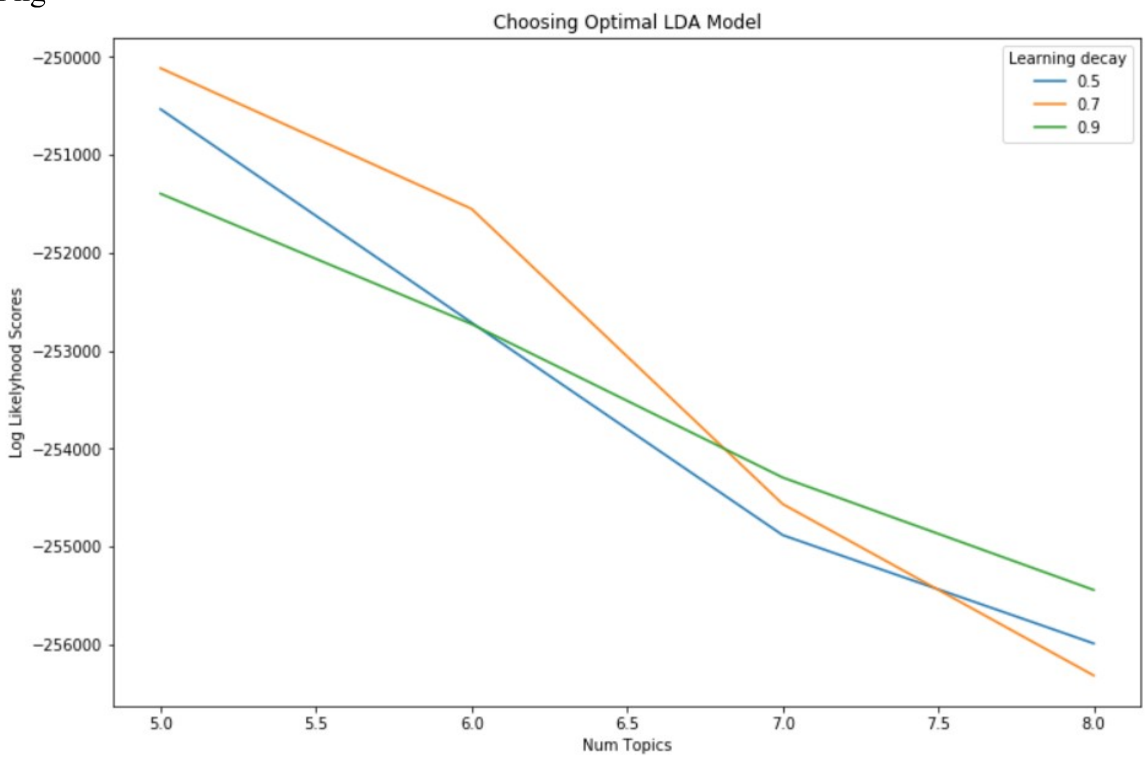

Figure 12: Learning Decay for 'Acupuncturists'.

```
# Apply Style
df_document_topics = df_document_topic.head(
    15).style.applymap(color_green).applymap(make_bold)
print(df_document_topic)

df_topic_distribution =
df_document_topic['dominant_topic'].value_counts(
).reset_index(name="Num Documents")
df_document_topic["docs_text"]=data
df_topic_distribution.columns = ['Topic Num', 'Num Documents']

print(df_topic_distribution)
df_document_topic.to_csv("output_docs_topic_"+ specialization + ".csv")
```

```
          Topic0  Topic1  Topic2  Topic3  Topic4  dominant_topic
Doc0        0.04    0.24    0.25    0.43    0.04               3
Doc1        0.69    0.00    0.00    0.19    0.11               0
Doc2        0.01    0.01    0.95    0.01    0.01               2
Doc3        0.76    0.01    0.12    0.01    0.11               0
Doc4        0.21    0.02    0.02    0.73    0.02               3
Doc5        0.35    0.01    0.08    0.56    0.01               3
Doc6        0.01    0.01    0.01    0.50    0.46               3
Doc7        0.04    0.04    0.04    0.84    0.04               3
```

Doc8	0.01	0.01	0.15	0.01	0.83	4
Doc9	0.50	0.01	0.01	0.16	0.32	0
Doc10	0.01	0.01	0.01	0.01	0.96	4
Doc11	0.07	0.07	0.07	0.73	0.07	3
Doc12	0.01	0.24	0.01	0.06	0.68	4
Doc13	0.01	0.01	0.01	0.46	0.52	4
Doc14	0.00	0.00	0.29	0.51	0.19	3
Doc15	0.01	0.01	0.67	0.30	0.01	2
Doc16	0.20	0.20	0.20	0.20	0.20	0
Doc17	0.48	0.02	0.02	0.02	0.46	0
Doc18	0.00	0.00	0.19	0.26	0.54	4
Doc19	0.01	0.01	0.01	0.01	0.97	4
Doc20	0.01	0.01	0.10	0.65	0.23	3
Doc21	0.02	0.02	0.19	0.38	0.39	4
Doc22	0.48	0.03	0.03	0.44	0.03	0
Doc23	0.01	0.01	0.01	0.83	0.13	3
Doc24	0.41	0.03	0.03	0.51	0.03	3
Doc25	0.01	0.01	0.01	0.39	0.57	4
Doc26	0.02	0.02	0.02	0.93	0.02	3
Doc27	0.02	0.02	0.02	0.94	0.02	3
Doc28	0.01	0.06	0.42	0.39	0.13	2
Doc29	0.02	0.02	0.29	0.40	0.27	3
...	...	...	...	...	...	...
Doc5605	0.14	0.01	0.17	0.01	0.67	4
Doc5606	0.28	0.01	0.15	0.38	0.18	3
Doc5607	0.01	0.01	0.01	0.94	0.01	3
Doc5608	0.10	0.01	0.19	0.09	0.62	4
Doc5609	0.11	0.01	0.01	0.28	0.58	4
Doc5610	0.03	0.03	0.03	0.90	0.03	3
Doc5611	0.02	0.02	0.02	0.94	0.02	3
Doc5612	0.12	0.01	0.30	0.24	0.32	4
Doc5613	0.09	0.07	0.27	0.01	0.55	4
Doc5614	0.01	0.01	0.42	0.01	0.54	4
Doc5615	0.45	0.03	0.03	0.44	0.03	0
Doc5616	0.30	0.03	0.62	0.03	0.03	2
Doc5617	0.22	0.01	0.01	0.56	0.20	3
Doc5618	0.38	0.01	0.01	0.58	0.01	3
Doc5619	0.43	0.03	0.03	0.48	0.03	3
Doc5620	0.02	0.09	0.41	0.46	0.02	3
Doc5621	0.51	0.02	0.02	0.44	0.02	0
Doc5622	0.44	0.02	0.02	0.25	0.28	0
Doc5623	0.03	0.03	0.03	0.88	0.03	3
Doc5624	0.26	0.02	0.02	0.02	0.68	4
Doc5625	0.12	0.12	0.02	0.73	0.02	3
Doc5626	0.80	0.05	0.05	0.05	0.05	0
Doc5627	0.73	0.07	0.07	0.07	0.07	0
Doc5628	0.10	0.10	0.60	0.10	0.10	2
Doc5629	0.20	0.20	0.20	0.20	0.20	0
Doc5630	0.06	0.00	0.01	0.46	0.47	4
Doc5631	0.01	0.04	0.66	0.01	0.28	2

```
Doc5632    0.13    0.01    0.77    0.08    0.01                2
Doc5633    0.01    0.01    0.48    0.50    0.01                3
Doc5634    0.20    0.20    0.20    0.20    0.20                0

[5635 rows x 6 columns]
   Topic Num  Num Documents
0          4           1693
1          3           1627
2          0           1051
3          2            935
4          1            329
```

```python
# Topic-Keyword Matrix
df_topic_keywords = pd.DataFrame(best_lda_model.components_)

# Assign Column and Index
df_topic_keywords.columns = vectorizer.get_feature_names()
df_topic_keywords.index = topicnames

# View
df_topic_keywords.head()

# Topic-Keyword Matrix
df_topic_keywords = pd.DataFrame(best_lda_model.components_)

# Assign Column and Index
df_topic_keywords.columns = vectorizer.get_feature_names()
df_topic_keywords.index = topicnames

# View
df_topic_keywords.head()


# Show top n keywords for each topic
def show_topics(lda_model, vectorizer=vectorizer, n_words=20):
    keywords = np.array(vectorizer.get_feature_names())
    topic_keywords = []
    for topic_weights in lda_model.components_:
        top_keyword_locs = (-topic_weights).argsort()[:n_words]
        topic_keywords.append(keywords.take(top_keyword_locs))
    return topic_keywords


# TODO put the number you want
topic_keywords = show_topics(
    vectorizer=vectorizer, lda_model=best_lda_model, n_words=15)

# Topic - Keywords Dataframe
df_topic_keywords = pd.DataFrame(topic_keywords)
```

```
df_topic_keywords.columns = ['Word ' + str(i)
                             for i in
range(df_topic_keywords.shape[1])]
df_topic_keywords.index = ['Topic ' + str(i)
                           for i in range(df_topic_keywords.shape[0])]


print(df_topic_keywords)

print("Save model.....")

joblib.dump(best_lda_model, 'best model'+specialization +
            '_'+str(best_lda_model.n_components)+'.pkl')
```

```
            Word 0     Word 1    Word 2        Word 3    Word 4     Word 5
\
Topic 0       help   medicine   chinese   acupuncture    health  treatment
Topic 1       elle       pour       son           une       amy        que
Topic 2       tell       feel      year         month      time        day
Topic 3  recommend       good      time        doctor  healthcare
consumer      great
Topic 4       pain  treatment      help   acupuncture      feel       year

            Word 6 Word 7      Word 8     Word 9 Word 10        Word 11
\
Topic 0     doctor    year    pregnant  recommend    issue         caring
Topic 1      blood     jai  traitement        tra      tre            qui
Topic 2        say    come    nathalie        tcm     week            bad
Topic 3       care    help        know     highly     make  knowledgeable
Topic 4  recommend  thank        treat    problem    issue           life

        Word 12        Word 13    Word 14
Topic 0     try           care  pregnancy
Topic 1     est           vous        pas
Topic 2   start           body       make
Topic 3    feel  acupuncturist  treatment
Topic 4    good         suffer       time
Save model.....
```

```
['best modelAcupuncturists_5.pkl']
```

```
pyLDAvis.enable_notebook()

panel = pyLDAvis.sklearn.prepare(
```

```
best_lda_model, data_vectorized, vectorizer, mds='tsne')
panel
```

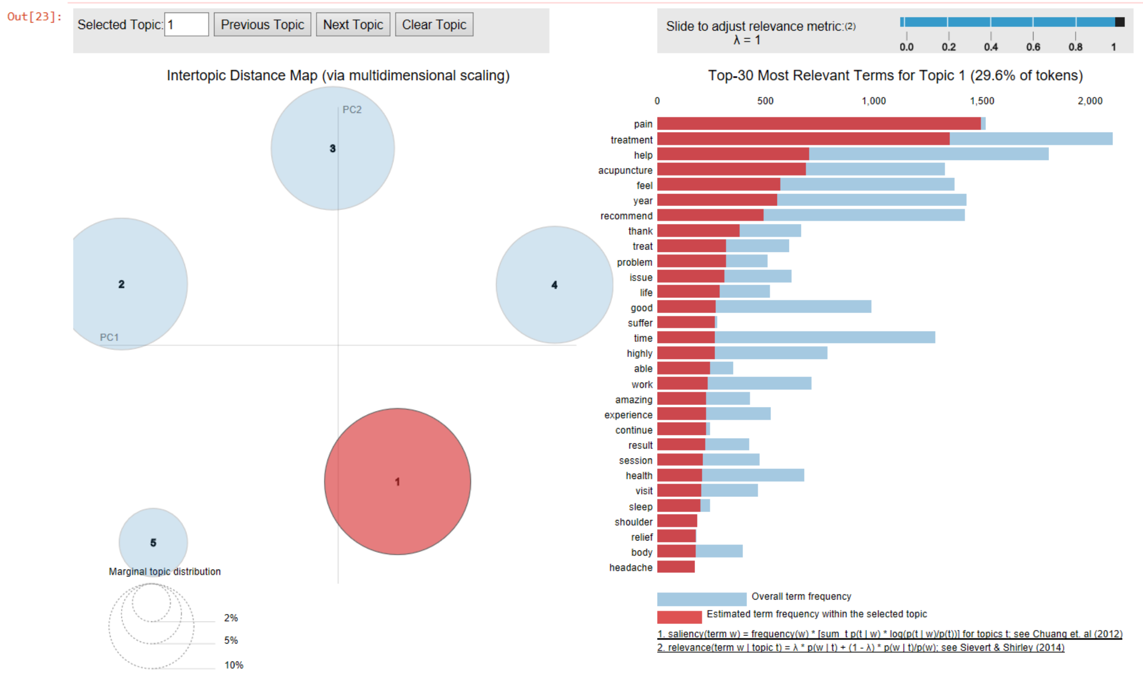

Figure 13: Number of Topics for ‘Acupuncturists’

APPENDIX C SPSS Workings

Alberta (AB)

	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	12473	4.1557	111.65996	.99980

One-Sample Test

	Test Value = 3.74					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Average_Ratings	-36.138	12471	.000	-.58401	-.6157	-.5523

British Columbia (BC)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	18188	2.9629	1.87625	.01391

One-Sample Test

	Test Value = 3.74					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Average_Ratings	-55.855	18187	.000	-.77706	-.8043	-.7498

Manitoba (MB)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	4548	3.0106	1.84759	.02740

One-Sample Test

	Test Value = 3.74					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Average_Ratings	-26.625	4547	.000	-.72942	-.7831	-.6757

New Brunswick (NB)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	2317	3.2637	1.79714	.03734

One-Sample Test

	Test Value = 3.74					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Average_Ratings	-12.758	2316	.000	-.47632	-.5495	-.4031

Newfoundland and Labrador (NFL)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	1309	3.3615	1.70348	.04708

One-Sample Test

	Test Value = 3.74					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Average_Ratings	-8.039	1308	.000	-.37853	-.4709	-.2862

Nova Scotia (NS)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	3503	3.1767	1.81770	.03071

One-Sample Test						
	Test Value = 3.74					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Average_Ratings	-18.341	3502	.000	-.56327	-.6235	-.5031

Northwest Territories/Yukon/Nunavut (NWT)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	169	2.3863	2.19804	.16908

One-Sample Test						
	Test Value = 3.74					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Average_Ratings	-8.006	168	.000	-1.35367	-1.6875	-1.0199

Ontario (ON)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	45312	2.8760	1.92643	.00905

One-Sample Test	
	Test Value = 3.74

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Average_Ratings	-95.471	45311	.000	-.86401	-.8817	-.8463

Prince Edward Island (PEI)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	432	2.9109	1.92607	.09267

One-Sample Test						
	Test Value = 3.74					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Average_Ratings	-8.947	431	.000	-.82910	-1.0112	-.6470

Quebec (QC)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	414	4.4394	.53678	.02638

One-Sample Test						
	Test Value = 3.74					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Average_Ratings	26.512	413	.000	.69942	.6476	.7513

Saskatchewan (SASK)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Average_Ratings	3632	3.2061	1.75027	.02904

One-Sample Test						
	Test Value = 3.74					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Average_Ratings	-18.383	3631	.000	-.53388	-.5908	-.4769

APPENDIX D LDA results on the dataset

This Appendix Dontains the LDA model results both for;

1) Healthcare providers Specialties
2) Geographic Locations (Provinces)

1) Healthcare providers Specialties

Table 9: Sample Comments from Specialties

Specialty	Category	Comments
Acupuncturist	**Skills**	"Dr. was informative, friendly, and helpful. She better understood my pain …"
	Pain	" Dr. took the time to help me with my back pain. He is very knowledgeable …"
	Knowledge	"She is extremely friendly and amazingly knowledgeable. She made me feel so comfortable…"
Cardiologist	**Manners**	"Knowledgeable, easy to speak with. Ive been going to him for awhile now and he knowledgeable…"
	Friendly	"Willing to spend the time and answer all my questions... friendly and personable…"
	Attention	"Great Doctor. Very helpful. I enjoyed the service and highly recommended to anyone…"
Dentist	**Anecdotal**	"Dr. and his staff are absolutely wonder people. Thank you for what you do…"
	Service	"Horrible wait times. Average cavity takes 3 hours. They sit you in the chair and…"
	Recommendation	"Great service. Caring and competent. The best! Highly recommended…"
Family Doctor / G.P.	**Compassion**	"I went to emergency and Dr. Richard was very helpful and thorough and here is a word to those worry about his bedside manner. Good Docs are focused on your health not PR. He did follow up for me until my Doctor in Brooks returns from holiday. I feel i was well cared for in the Bassano emergency!"

	Schedule	“When you do get an appointment Dr. Adams is on time which is refreshing however dont expect to get an appointment when youre in extreme pain and need a doctor - like your family doctor. Dont even think about leaving a message on the answering machine because no one calls you back. …”
	Friendly	“Cette docteure est un danger pour les patients. Les commentaires sont soit trÃƒÂ¨s positifs ou trÃƒÂ¨s nÃƒÂ©gatifs. Elle est instable et irrÃƒÂ©guliÃƒÂ¨re. Elle ne fait pas les suivis nÃƒÂ©cessaires.”
Chiropractor	**Staff**	“suffered long time from stroke. great person and teacher. thank you.”
	Friendly	“Le meilleur mÃƒÂ©decin de famille que lon peut avoir!”
	Helpfulness	“Very knowledge and helpfulness dentist, with strong work ethic.”
Plastic / Cosmetic Surgeon, Physician	**Procedure**	“Very pleased with entire breast augmentation experience. Found Dr Niessen & his entire staff to be very friendly and comforting throughout every visit. Could not be happier with the results!”
	Schedule	“DonÃ¢Â€Â™t waste youÃ¢Â€Â™re time and money......35 year old male had tummy tuck procedure done not once but twice left with deformity below belly button....prior to the surgery told me to relax that he had done hundreds of tummy tuck....well Dr. Niessen think you have more to go....horrible doctor.....donÃ¢Â€Â™t be fooled.”
	Positive	“Surgery was cancelled with no details as to why. Seems to be something serious going on with this practice.”
Gynecologist (OBGYN)	**Interval**	“I did not care for her dismissive manner. I only met her for 5 minutes (after waiting over an hour past my appointment time). I am disappointed she did not take the time because she was running late, and just gave me pamphlets about IUDs. Just make another appointment if I had questions.”
	Advise	“Dr. has been nothing but amazing. My family doctor kept dismissing my pelvic pain for months, making me jump through hoops, and then I had to beg him to refer me to see her. He said that gynaecologists hate getting referrals for "pelvic pain" because they are so vague. I knew something wasnt right and wanted answer. Dr. Kenyon listened to my issues and concerns and addressed them accordingly. When my issues didnt improve, she suggested laproscopic surgery to investigate….”
	Perception	“I dont know why every one is giving dr. a bad review, when I had my surgery to have a ovarian cyst removed she was actually very friendly and kind, people who dont have anything nice to say shouldnt say anything at all, I feel bad that this world is filled with negative comments. “
General Surgeon	**Expert**	“My daughter and my father both were operated by Dr. Hutchinson both with excellent results. Highly reccomended. My son has been operated after trauma by Dr. with somewat diffrent results we are going to have this procedure re-done at

		Sunnybrook.”
	Care	“This doctor was fabulous for both my husband who cut the tip of his finger off and had surgery to shorten the bone and reconstruct the skin around it and my son who broke his hand and after the Cobourg hospital set it (STILL BROKEN) he reset and had put pins in to keep it set (odd location of break) as a surgery. He was very helpful, knowledgable and patient with all my questions. …”
	Disposition	“Ce chirurgien fait de lexcellent travail mais la dite secrÃƒÂ©taire ne mÃƒÂ©rite pas cette appellation.. La santÃƒÂ© mentale de cette femme est dÃƒÂ©faillante Ãƒ tout point de vue. …”
Optometrist	**Recommend**	“I came to Dr. when one eye dropped and she told me it was just cosmetic and turned me away. I had to go to Edmonton to get proper care.”
	Professionalism	“An excellent optometrist. I have retired and moved to Sherwood Park, but after a bad experience with a new optometrist, I will be driving 5 hours to see her for future eye exams.”
	Compassion	“I found her to be very helpful. She helped me understand my problem and I would encourage anyone who asks to take an appointment with her. Excellent service and very helpful.”
Orthopedic Surgeon	**Immense**	“I had a total knee replacement done by Dr. sixteen months ago. I couldnÃ¢Â€Â™t be more pleased with the results! Dr. Maragh took time to explain proÃ¢Â€Â™s and conÃ¢Â€Â™s unique to my situation due to previous orthopaedic surgeries and possible alternatives….”
	Torment	“Dr. operated on my dads bunion on his right foot last year. He left the dressing on my dads foot for one month before he would see him again. My dad thought that he had put a cast on his foot because it felt so hard but it was actually that the dressing had hardened with dried blood. …”
	Schedule	“Today is 1 year post-surgery to remove a neuro-fibroma on the top of my foot. Since such a large area of my foot was involved, I was afraid I would end up with complications or needing a skin graft. Dr. Bridge did a fantastic job, and I also received excellent care from the KGH wound specialist….”
Pediatrician	**Considerable**	“wonderful help you are receiving, instead of complaining think of those beautiful babies that didnt make it. You should be thankful for your wonderful child. Talk to some of the wonderful parents whose children did not make it….”
	Staff	“My son was born at 26 weeks. Maybe Dr.Saigal havent spoke loud enough but her knoweledge helped my son to be alive today. Thank you doctor. Tatyana Krmpotic”
	Friendly	“We love Dr! Shes so friendly, knowledgeable, and wonderful with kids. She makes them feel important and reassures them so well. Yes, sometimes theres a wait to see her but that is clearly documented on the info you receive prior to your appointment. Any specialist has a wait- and at least this place has an awesome

		waiting room :)”
Orthodontist	**Helpfulness**	“My son really liked Dr. He was great with my son. He pushed for jaw surgery, and in the end, we went with it, and even though it was a tough recovery, my son is glad he decided to do it.”
	Procedure	“FANTASTIC orthodontist! Best in Saskatoon -- I had been to other clinics for consultations, but Dr. Kurz was the only one I felt comfortable with. Enthusiastic, …”
	Schedule	“I believe that Dr. is extremely knowledgable, helpful, and above all, very caring. I appreciate that he takes into consideration the expense of orthodontics and is sensitive to that. I trust him completely with my daughter and her care.”
Psychiatrist	**Positive**	“An encounter with Dr. has touched my life as would Saint Michael the Archangel.”
	Staff	“this doctor literally saved my sisters life during a mental health crisis. I have no words to say how much I appreciate her”
	Helpfulness	“Love this dr. Always takes time to listen. I think who writes negative about her are those who donÃ¢Â€Â™t get benzos or disability paperwork filled. Easy to blame dr rather then take responsibility of your actions and treatment.”
Ophthalmologist	**Schedule**	“I feel very confident and comfortable with Dr. Yearsley. He gives you all the facts in a straightforward, easy to understand manner and is an extremely pleasant man.”
	Procedure	“Dr. performed Cataract surgery on each eye three months apart . The surgeries reduced the need for glasses to reading glasses only. Dr. Yearsly was very informative and supportive both pre and post surgery. He is a consummate professiona. It is no surprise that he is very highly rated.”
	Staff	“My first visit got off to a bad start. Staff seemed somewhat unprofessional: whispering conversations interspersed with laughter; not paying attention to making sure the patients were treated in proper order. I was told the appointment would take an hour but it took two. Not the doctors fault. He was very professional and thorough.”
Oncologist / Hematologist	**Compassion**	“Excellent Dr. Excellent manners. He was made for his profession.”
	Skills	“A well-mannered real man of class , great listener and shows respect. Has patience and is knowledgeable. I rate him highly professionally and as a human being.”
	Schedule	“Lucky here as we can get a second opinion at any time. It is great to be Canadian.”
Naturopath	**Recommend**	“I highly recommend Dr. Habert. She is helpful and knowledgeable and I always feel like I can have a conversation about my health with her. Dr. has helped me to feel so much better!”

	Schedule	“She is very knowledgeable, has a calming demeanor, and is easy to talk to. If you are a patient of hers, you are truly in good hands! The compassion and understanding that I continually feel from her is extremely genuine. She is so supportive. I trust her with my health 100% from diet changes and natural supplement.”
	Helpfulness	“Dr. was the answer I have been looking for all along. I have been working closely with her for a year now and have never felt better. She definitely has the knowledge, stays current with research, and genuinely cares about making a difference in her patients lives…”
Emergency Room Doctor	**Professionalism**	“Dr. was my GP before making the move to the ER. He is fantastic! I would feel very comforted if I ever needed to go to the ER and he was on duty.”
	Knowledgeable	“Excellent docteur, tres efficace! Tres professionnel, on voit quil aime son travail”
	Positive	“Dr. from the Maisonneuve clinic is quick in diagnosing - which isnt necessarily what you want when dealing with your health, should the diagnosis be wrong. I got the impression that he wasnt really listening to me when I was telling him how I felt...”
Internist / Geriatrician	**Staff**	“He was very professional and spend time to evaluate my case in detail and didnÃ¢Â€Â™t dismiss and ignore any of the details.”
	Helpfulness	“Doctor is a vampire who kills patients by draining them of their blood! Stay away from her, get your loved ones out of her care and do not go to William Osler Hospital!!!”
	Schedule	“Dr. may be in excellent internist, but I myself have found that his conclusion or findings of a matter based on information from several sources unverified is concerning. …”
Podiatrist	**Recommend**	“I did not find that it was about getting me better but more about selling me stuff.”
	Procedure	“Made an appointment with Dr.Yuen and was brushed off to a chiropodist who works in the same office even though I was specific about booking the appointment with him. …”
	Schedule	“Goes above and beyond the call of duty. A very kind, caring man. Follows up on any condition requiring it. Found him punctual.”
Urologist	**Doctor**	“Very, very mediocre surgeon, but liked money. Did an operation to me regarding prostate cancer “
	Schedule	“He is an absolutely amazing doctor; thorough, knowledgeable, honest, and I could go on. His staff are friendly, courteous, professional. Thank you for being you Doctor Bella”
	Appointment	“This doctor shows the utmost respect for his patients as human beings. I cannot thank Dr Bella enough. He just saw me yesterday, on a Saturday, very late in the day after starting seeing his patients at 530 am. Dr. understands how much things hit us as men when things dont work. …”

Psychologist	**Helpfulness**	"I had the privlidge of being treated by Dr. Kamps about 9 years ago when I was 13 and until I finished high school in 2012, and to this day I have nothing but good things to say about her. Dr. is an amazingly kind, caring, and compassionate woman. …"
	Prescribe	"This supposed mental health doctor is paid by insurance companies to provide false reports on people that are truly injured. Although during an assessment he seems kind and actually is empathetic the reports he writes and decisions he makes are false. …"
	Understanding	"Do not hire him to prepare a section 211. I trusted the system to work as it should. I was wrong. As he stated, "sometimes I get it right, sometimes I get it wrong". For the sake of the children involved and for all of the experience he supposedly has, he should be getting it right."

In addition, figures 10, 12, 14, and 16 represent the visualization of the topics obtained through pyLDAvis. The characteristic of a good topic model is that it will have non-overlapping, large blobs representing each topic. This seems to be the case for the model output.

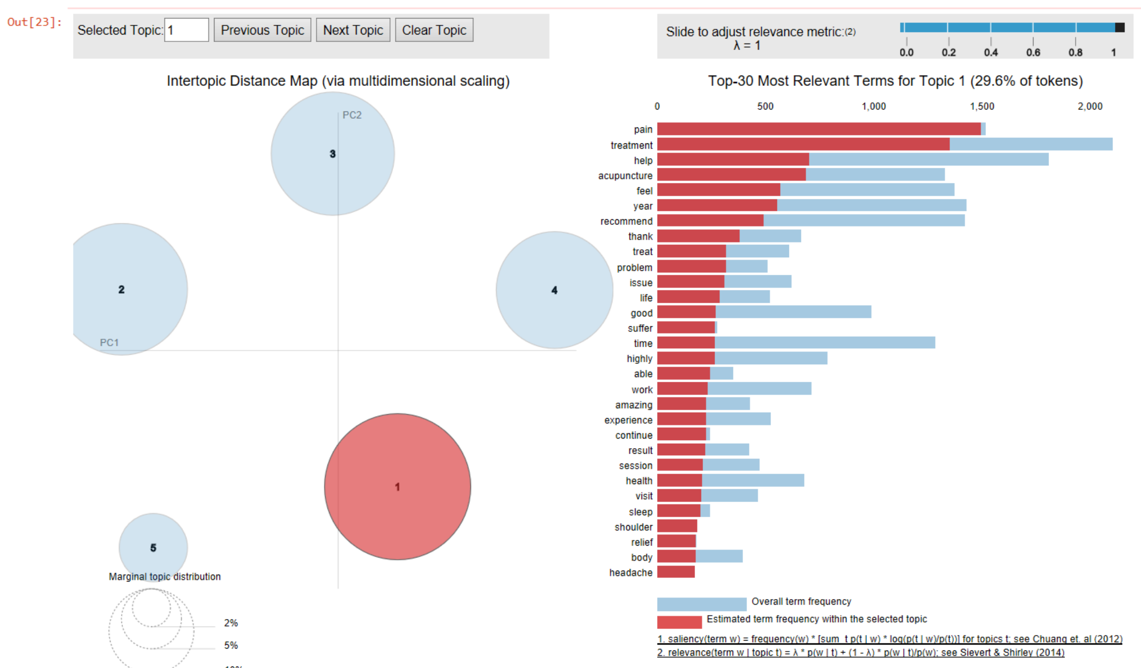

Figure 14: Number of Topics for 'Acupuncturists'

Figure 14 depicts the learning decay for Accupuncturists. As the output plot the log-likelyhood scores against num_topics, the plot shows number of topics i.e. 5, 6, 7, 8.

Furthermore, topic 7 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

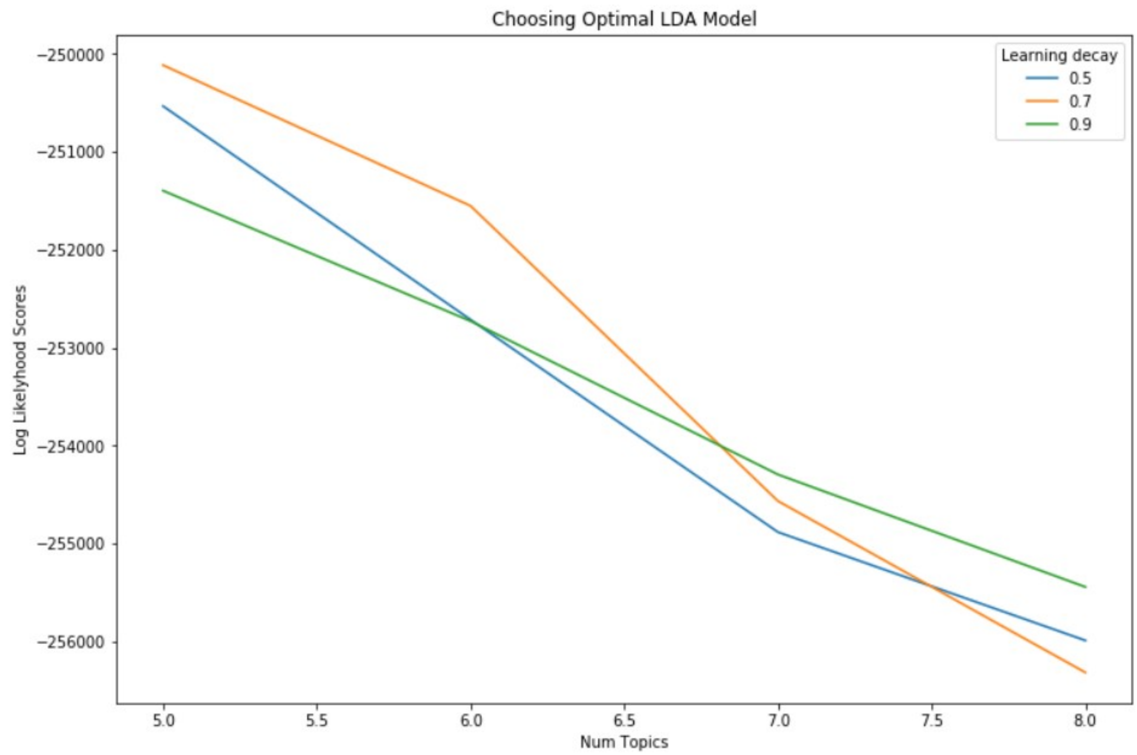

Figure 15: Learning Decay for 'Acupuncturists'.

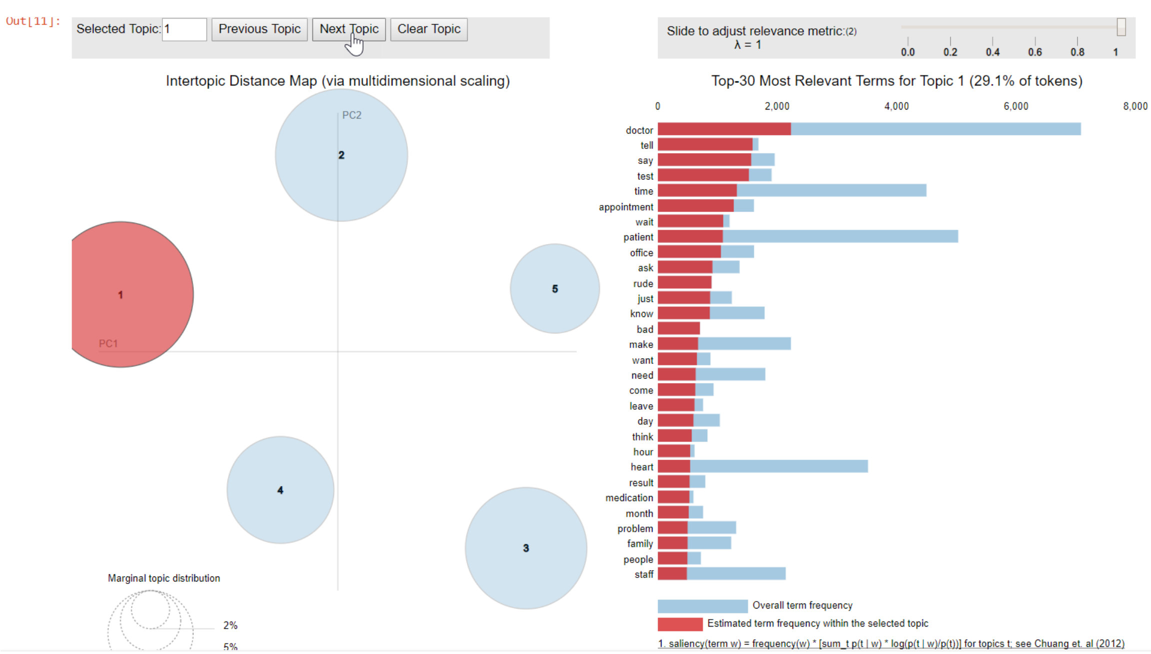

Figure 16: Number of Topics for 'Cardiologists'

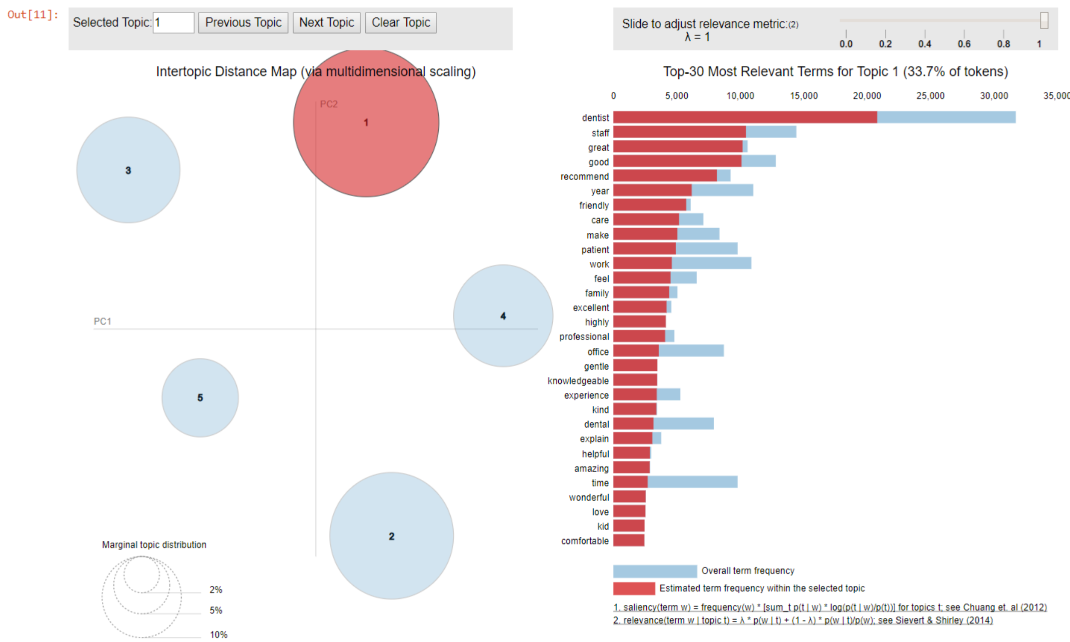

Figure 17: Learning Decay for Cardiologists.

Figure 17 depicts the learning decay for Cardiologists. As the output plot the log-likelyhood scores against num_topics. Similar to figure 11, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 7 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

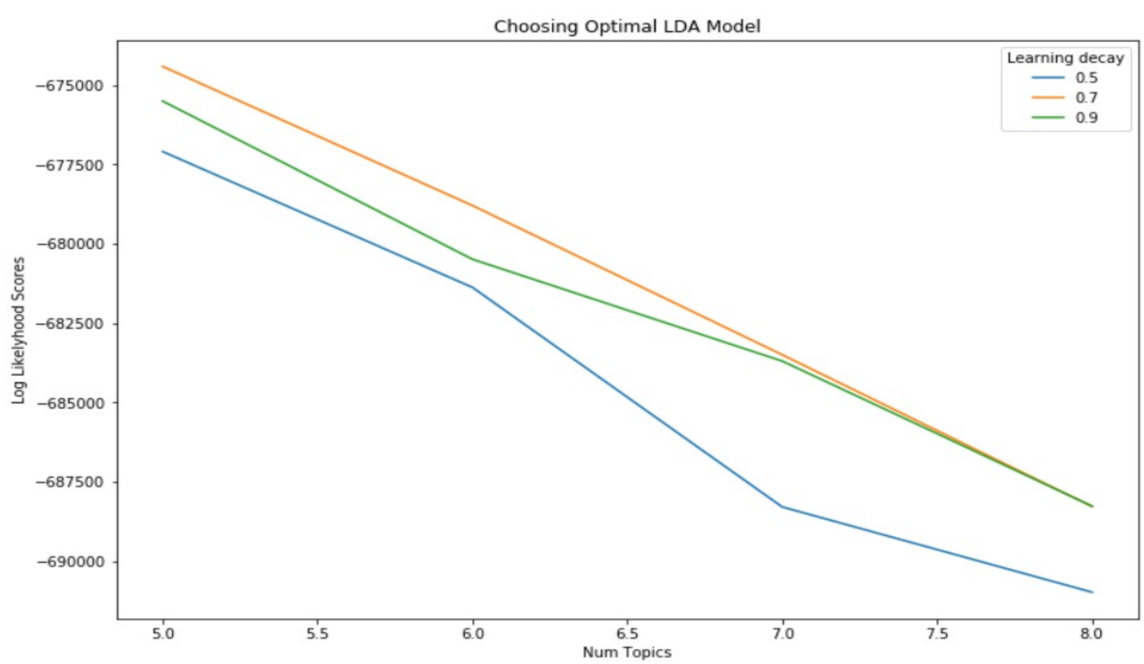

Figure 18: Number of Topics for 'Dentists'.

Figure 19 depicts the learning decay for Dentists. As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 7 has a relatively better performance score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

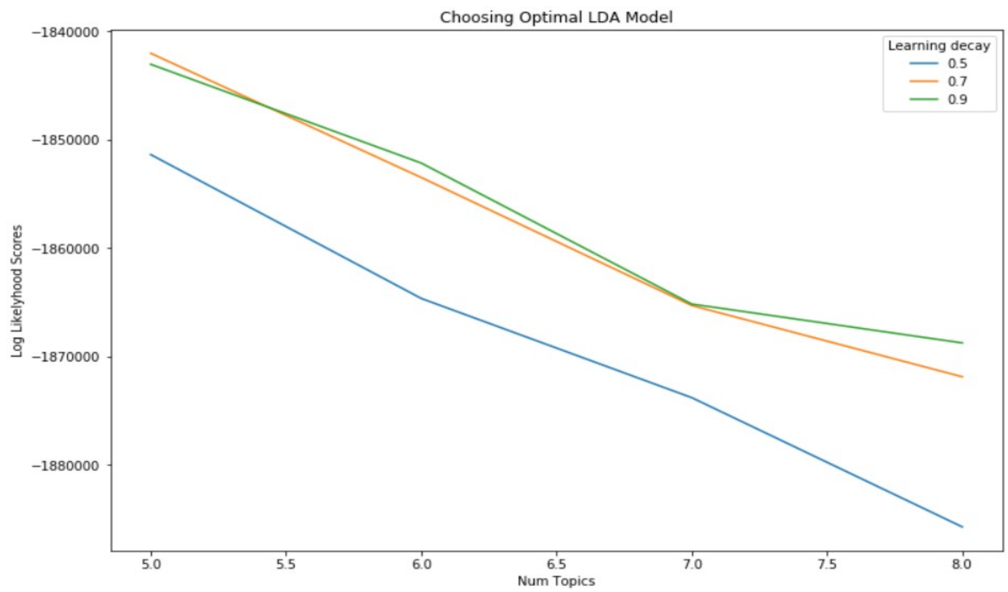

Figure 19: Learning Decay for ‘Dentists’.

Table 9 shows the category, most frequent LDA words, and a (possible) sample sentence taken from the output of the model. The categories in the first row represents document 8, with probabilities of 0.03 for topic 0, 0.03 for topic 1, 0.24 for topic 2, 0.03 for topic 3, and 0.66 for topic 4, and an overall highest score of 4 for the dominant topic. While, the second row represents document 10, with probabilities of 0.9 for topic 0, 0.03 for topic 1, 0.03 for topic 2, 0.03 for topic 3, and 0.03 for topic 4 with an overall highest score of 0 for the dominant topic. Furthermore, row three represents document 21, with probabilities of 0.04, 0.86, 0.03, 0.03, 0.03, for topic 0, 1, 2, 3, 4 respectively, with an overall highest score of 1 for the dominant topic.

2) Geographic Locations (Provinces)

Table 10: Sample Comments as per Provinces

Provinces	Category	Comments
Alberta (AB)	**Excitement**	"Dr. is U of As finest Grad of the Dentistry/Medicine Program. He knows his "craft" to the "T" (hence the T middle name). Too bad he wasnt my medical doctor. Couldve saved Alberta Health and Wellness from going bankrupt with his quick diagnoses, first rate prognosis, and clinic wait times. Too bad he decided not to take Medicine. Cancer wouldve been eradicated sooner. Oh well, lucky for us dental patients!!LOL!!"
	Satisfaction	"Very excellent doctor, gives all the options and does not try to push any work, very thorough and honest, i knw i can always count on getting the best quality care."
	Friendly	"I just felt that my child did not need a root canal on a baby tooth and did not realize that was what was going on until the appointment took over an hour. I find the staff friendly to patients, but too pushy for your insurance information, and not very nice to each other. "
Ontario (ON)	**Appointment**	"Dr. was informative, knowledgeable,accomodating with appointment times and overal polite and helpful."
	Recommendation	"I have been to a few ENTs, but Dr. was by far the best. She was on time, listened to my questions and made an effort to help. If you get Dr. Osborne youre in good hands!"
	Professionalism	"Doctor discovered my cancer. She led me in the right direction to get the problem attacked and under control. I went to MD Anderson cancer center in Houston, Tx for my very unique surgery requirements, but I will never forget the

		comprehensive physical exam that Annie gave me and.... gives me every year, and how important that was in my health as it stand today. Doctor Hum is "top shelf" in my books. Richard Cohen.”
British Columbia (BC)	**Amazing**	“Dr. is an amazing doctor. He is kind and patient with his patients. If your in Vernon and happen to be in the ER.. Dr Bunten is your man. Amazing man and very through. You will not be disappointed.”
	Service	“Dr. is an outstanding human being. He is caring, patient and intelligent. After 15 years in his care, and also the care of my late husband, I have recommended him to my valued friends.”
	Thorough	“He was very professional and spend time to evaluate my case in detail and didnÃ¢Â€Â™t dismiss and ignore any of the details.”
Manitoba (MB)	**Great**	“He is one of the best chiropractors in the city. My family a d Ive seen him for many years. ”
	Patient	“I have been a patient of Dr. for almost 13 years, hes helped with chronic pain and issues resulting from several car accidents. The holistic approach that he uses is why I continue coming to Markham Chiropractic. ”
	Professionalism	“Worst doctor. dosent pay attention to the patients medical history .she gave me medicine according to her own preference.”
Nova Scotia (NS)	**Professionalism**	“Dr. took all four of my wisdom teeth out. His nurse was amazing and comple…”
	Service	“Dr. is always 30-60 minutes late, she only works part time, her appointme…”
	Recommendation	“Knowledgeable, easy to speak with. Ive been going to him for awhile now and he…”
New Brunswick (NB)	**Attitude**	“This dude needs an attitude adjustment. He has a serious "God complex" issue.”
	Doctor	“I had an emergency appt Dec 11/17 with Dr.

		Blacquiere and found him very concerned, very thorough and a very caring person. Both my husband and I are very impressed with his knowledge and handling of me. Thank you and please keep up the good work.”
	Appointment	“This is the 2nd time Ive seen Dr. Dylan Blacquiere. The first time, I was 5 minutes late and he was going to refuse to see me after I drove 1.5 hours to see him. He did see me and was short with me like I was wasting his time. After waiting almost a year for two mris, I seen him again today. This time he was 40 minutes late, did not apologize, and hurried me through the appointment. He interrupted me every single time that I tried to speak, talked down to me like I was a child, called me a liar, acted disrespectful and arrogant, and I wasnt even able to tell him about my new symptoms as he was basically pushing me out the door. He told me that Id need another doctor referral to see him again, and I assured him that I would NEVER see him again. Then I told everyone in the waiting room that hes an a-hole and cried all of the way home. Ive been sick for over 3 years, I dont need this! He needs to go take a course in social skills.”
Prince Edward Island (PEI)	**Recommendation**	“I have to say that Dr. saved my husbands life and still continues to do so. She makes sure that we are both taken care of and fights for what she believes in which are peoples lives. She continues to make sure she fights for her patients needs and wants. She is a fighter and that might be a trait that people may not like but if she is your physician that is a trait that you want because I know we do. Keep it going Dr. T. Thumbs up to the new receptionist you are doing an amazing job. ”
	Care	“Dr. is a very busy doctor with her call schedule she

		is away a lot which would make her receptionist job extremely difficult. Amy you are doing a great job dont let others bring ya down! You go girl!!!"
	Staff	"I know Dr. gives excellent care but whether patients know it or not, there is a lot more work behind the scenes for the secretarial staff than people realize. Its a chaotic environment and not an easy job to do. Amy keep your head held high and keep doing the very best you can!"
NewFoundland&Labrador (NL)	**Recommendation**	"Dr. is knowledgeable, compassionate and committed to helping his patients. Ive been seeing him for several years and never feel better than I do after a treatment. He explains things thoroughly, is super professional and also has a sense of humour. Best chiropractor in town, in my opinion. "
	Scheduling	"Dr. was filling in for my regular OB while he was away and I was in to see her twice during that time. This is my first pregnancy (through IVF), so I had a lot of questions and concerns. Dr. Ferguson is very warm and friendly and made me feel at ease right away. She is very knowledgeable and responded to all of my questions with great detail and explained everything thoroughly. I had quite a few questions at both of my appointments, but I never felt rushed. By the end of my appointments, I felt much more at ease and less anxious about my pregnancy. She gave me tons of reassurance. I hope to get to see her again during my pregnancy as she was absolutely amazing!"
	Humane	"Great doctor, very nice and understanding. Helped make me feel more comfortable before, during and after my section. Answered all my questions plus more. Friendly and kind. Made sure that me and baby were fine, always went above and beyond. Highly reccomend. "

Saskachewan (SASK)	**Knowledgeable**	"The best! After 24 years with epilepsy, brain surgery etc. I have a ton of confidence and respect for him. Hes been my epi for a year and a half in BC, we are very lucky to have him. Totally get why Sask patients love him... we do here too he is one of, if not the, most knowledgeable and kind doctors Ive ever met. Hes respectful when it comes to issues with medication etc., speaks his mind but gently. Anyone who became his patient when he joined the Epilepsy Clinic at VGH is very lucky and well looked after."
	Recommendation	"I think you will be very happy going to see Dr. Jason. He is funny and nice and treats people like they are his friends. I recommend him to whoever needs a dentist!"
	Wit	"……is a amazi g dentist! Funny, professional & helps you feel very comfortable."
Quebec (QC)	**Que**	"Je suis dÃƒÂ©solÃƒÂ© apprendre que tu revient plus Ãƒ la clinique de st Henri de levis vous savez je."
	Experience	"Dr. has been great. The staff was amazing. The old receptionist i knrw was terrific.The new reception is eithet overloaded or incompetent. They were supposed to send out panorex to specialist. They were asked weeks before by the Dr, the specialist and myself. The panorex were never received by the specialist. I had to repay because of this."
	Recommendation	"Highly recommended! Great team!!! Exceptional staff. The assistants always greet you with a smile. No one likes changed but in this case the new receptionists are great. Helpfull, considerate and trying to make the visit more pleasant. To the whole staff...... Thank You."

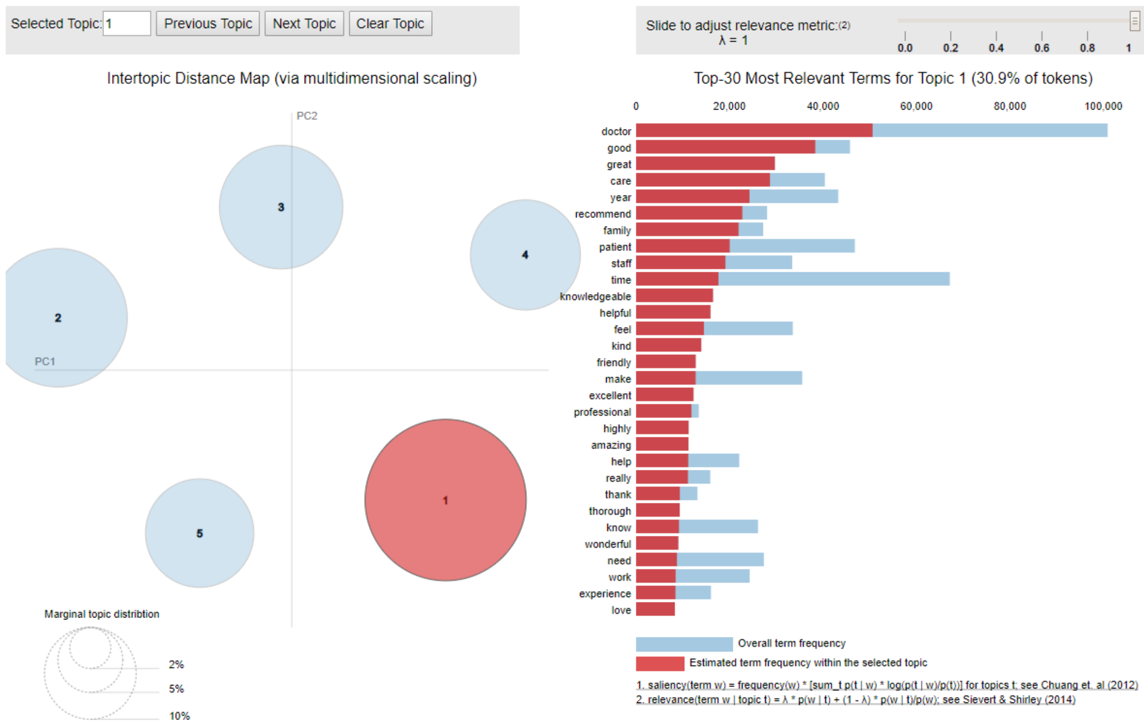

Figure 20: Number of Topics for 'Doctors in Alberta (AB)'.

Figure 21 depicts the learning decay for Doctors in Alberta (AB). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 7 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

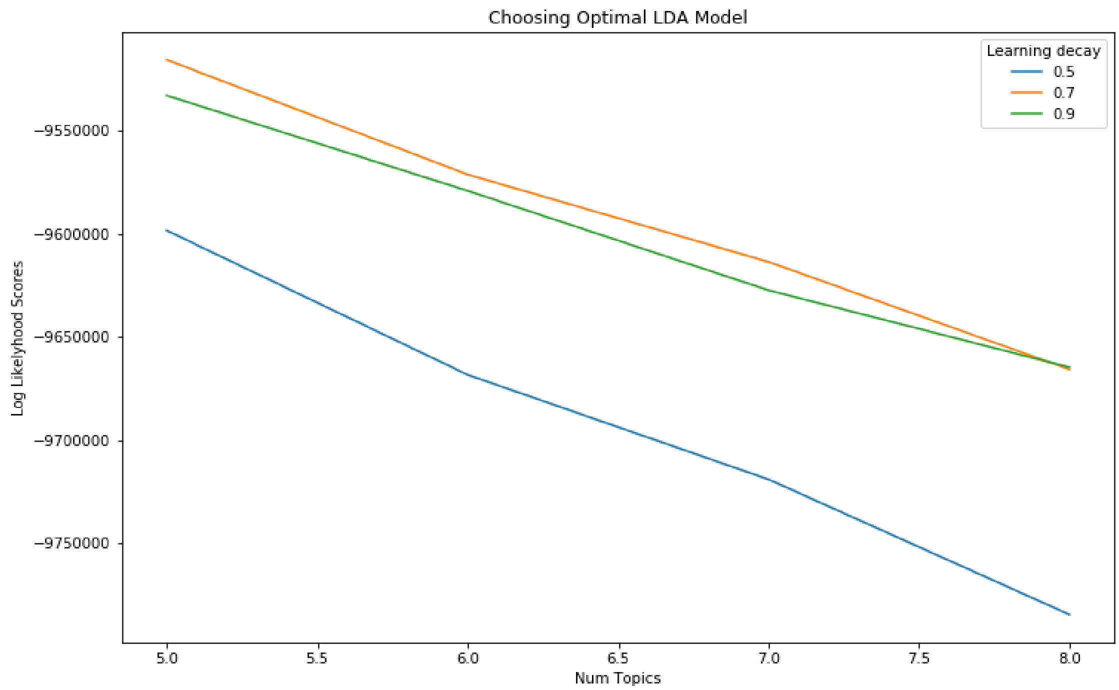

Figure 21: Learning Decay for ‘Doctors in Alberta (AB)’.

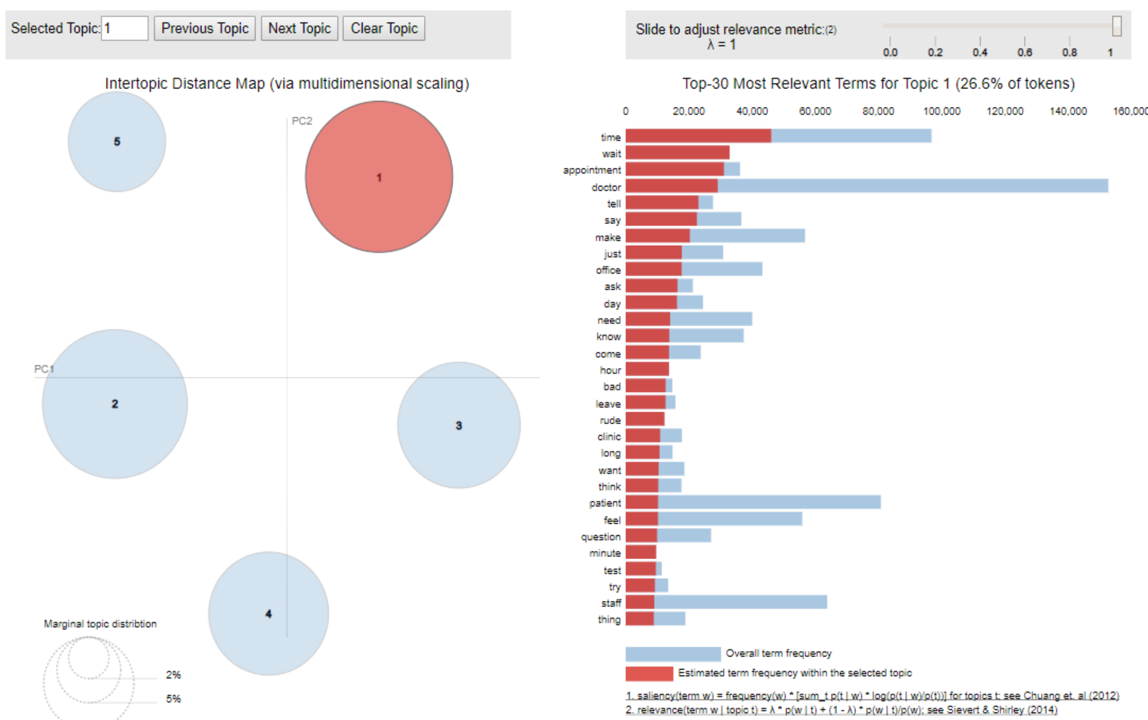

Figure 22: Number of Topics for ‘Doctors in Ontario (ON)’.

Figure 23 depicts the learning decay for Doctors in Ontario (ON). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 7 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

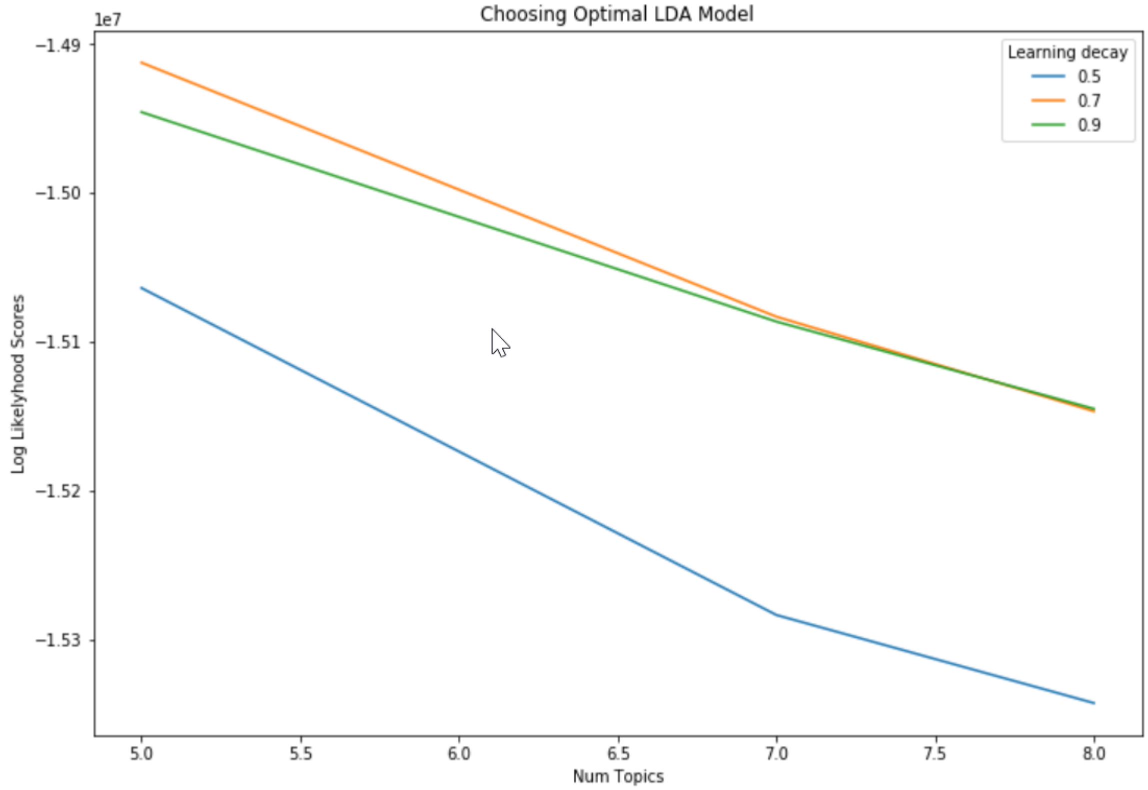

Figure 23: Learning Decay for 'Doctors in Ontario (ON)'.

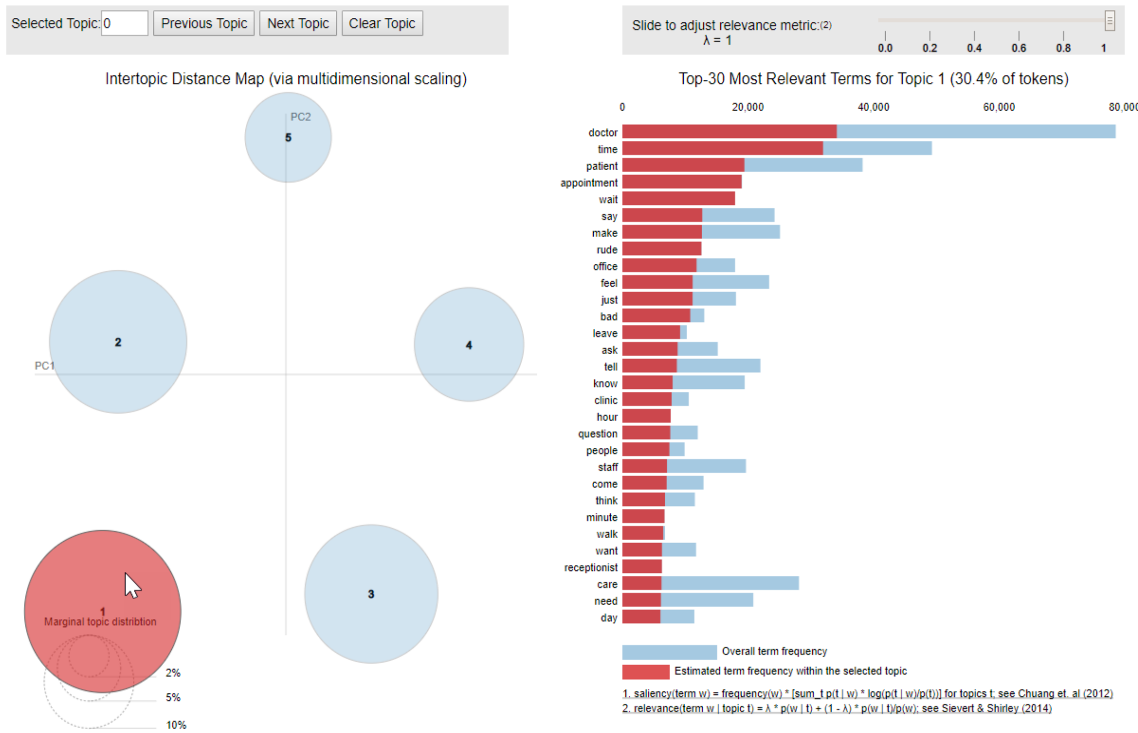

Figure 24: Number of Topics for 'Doctors in British Columbia (BC)'.

Figure 25 depicts the learning decay for Doctors in British Columbia (BC). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 7 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

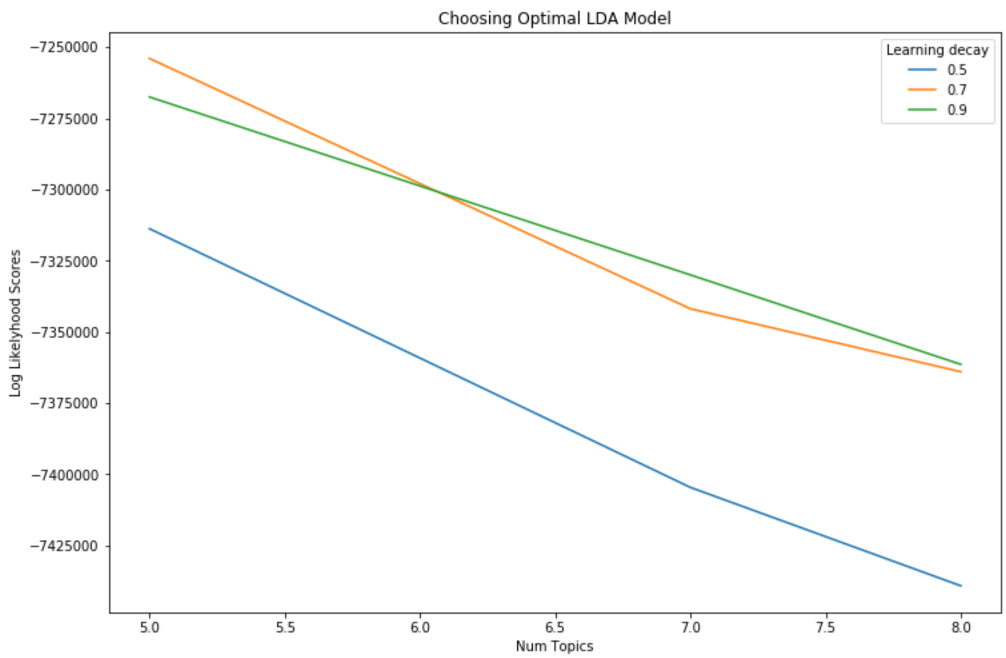

Figure 25: Learning Decay for 'Doctors in British Columbia (BC)'.

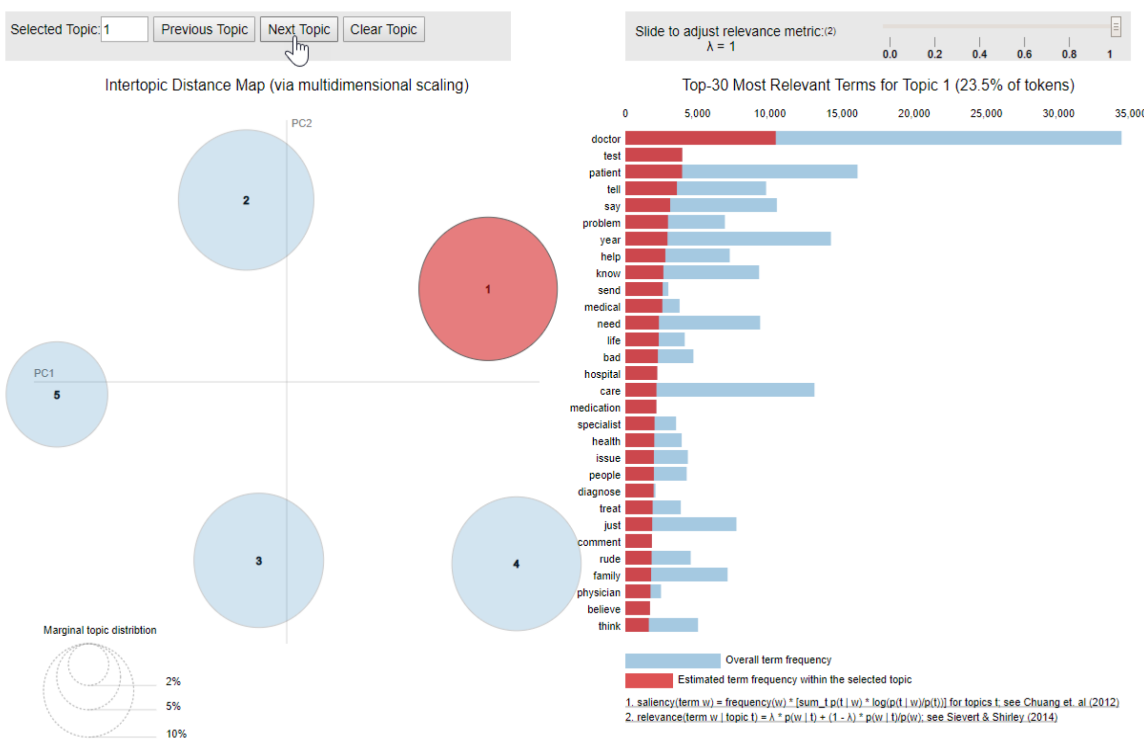

Figure 26: Number of Topics for 'Doctors in Manitoba (MB)'.

Figure 29 depicts the learning decay for Doctors in Manitoba (MB). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, both topic 7 and 9 have a better score. In addition, the `learning_decay` of 0.7 and 0.9 outperform 0.5.

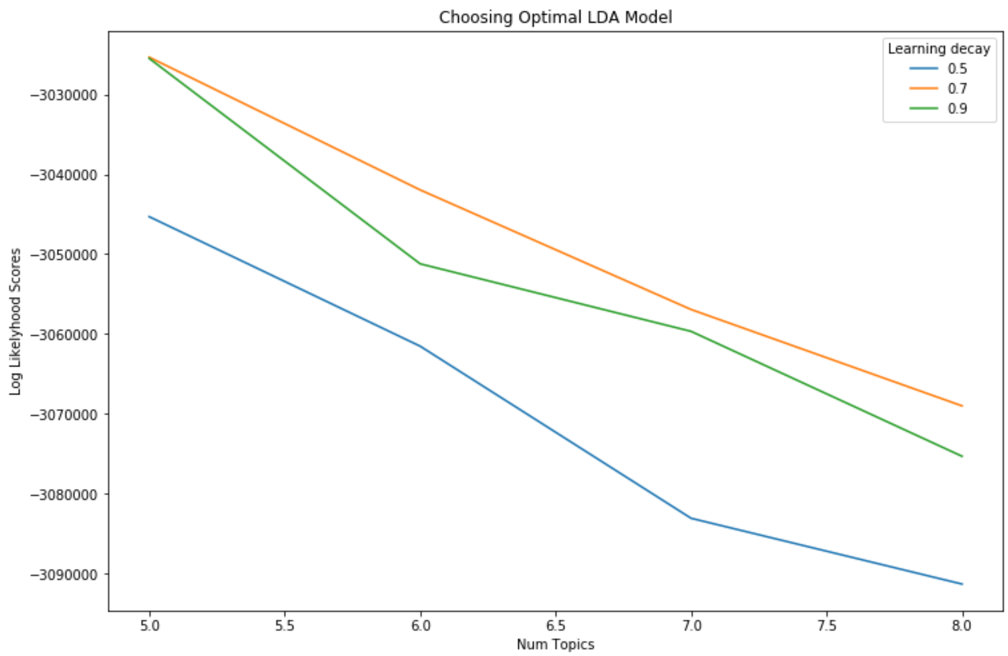

Figure 27: Learning Decay for 'Doctors in Manitoba (MB)'.

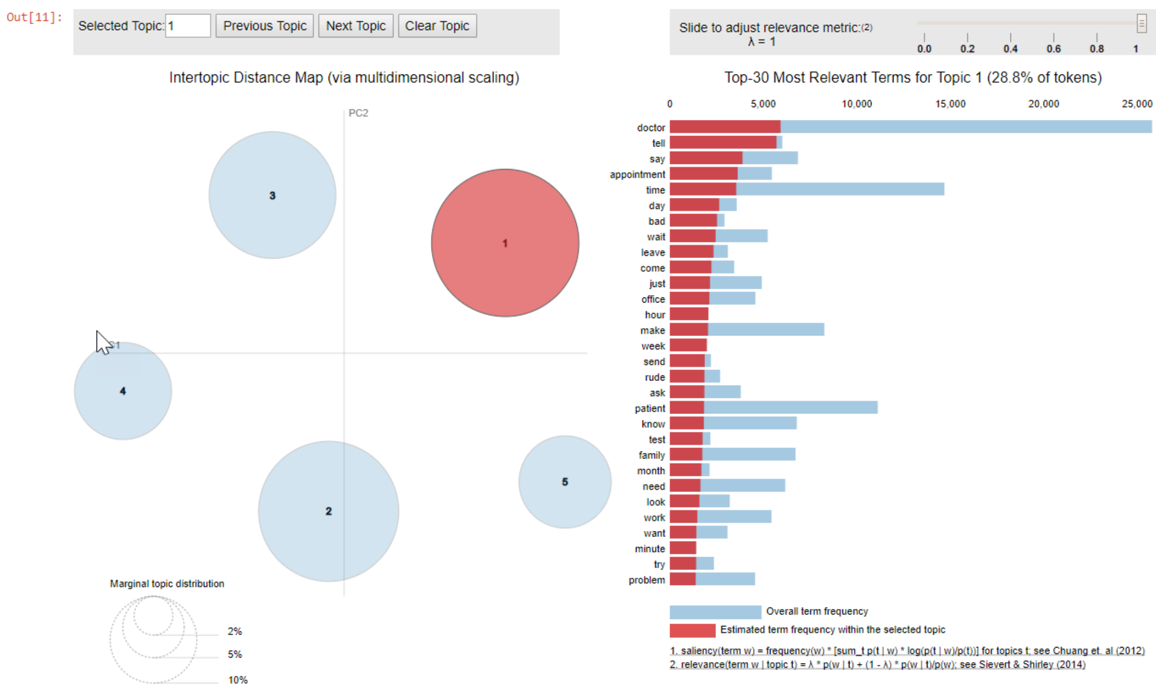

Figure 28: Number of Topics for ‘Doctors in Nova Scotia (NS)’.

Figure 29 depicts the learning decay for Doctors in Nova Scotia (NS). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

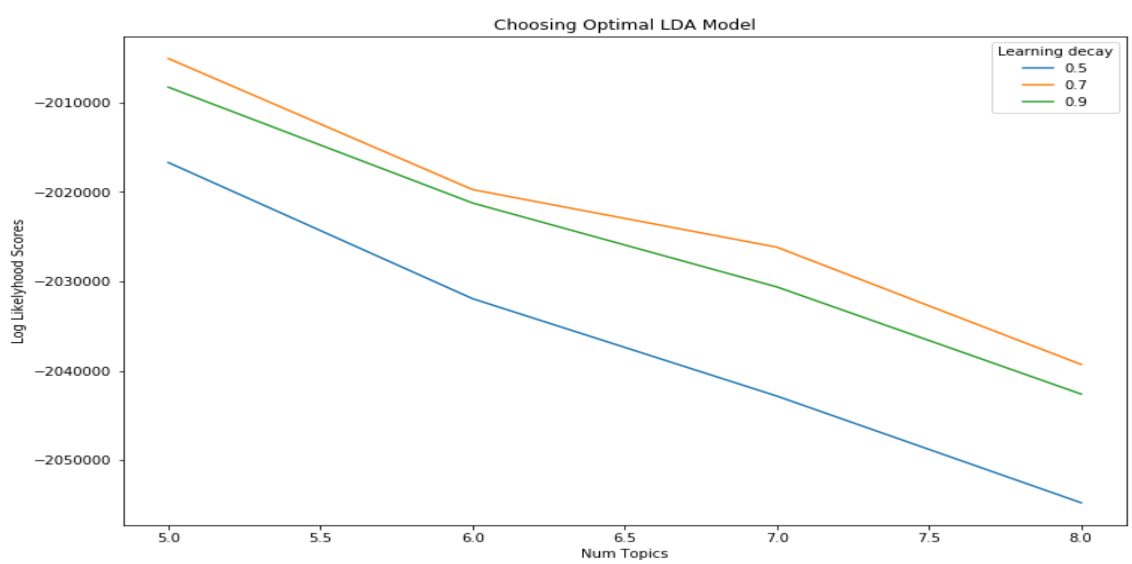

Figure 29: Learning Decay for ‘Doctors in Nova Scotia (NS)’.

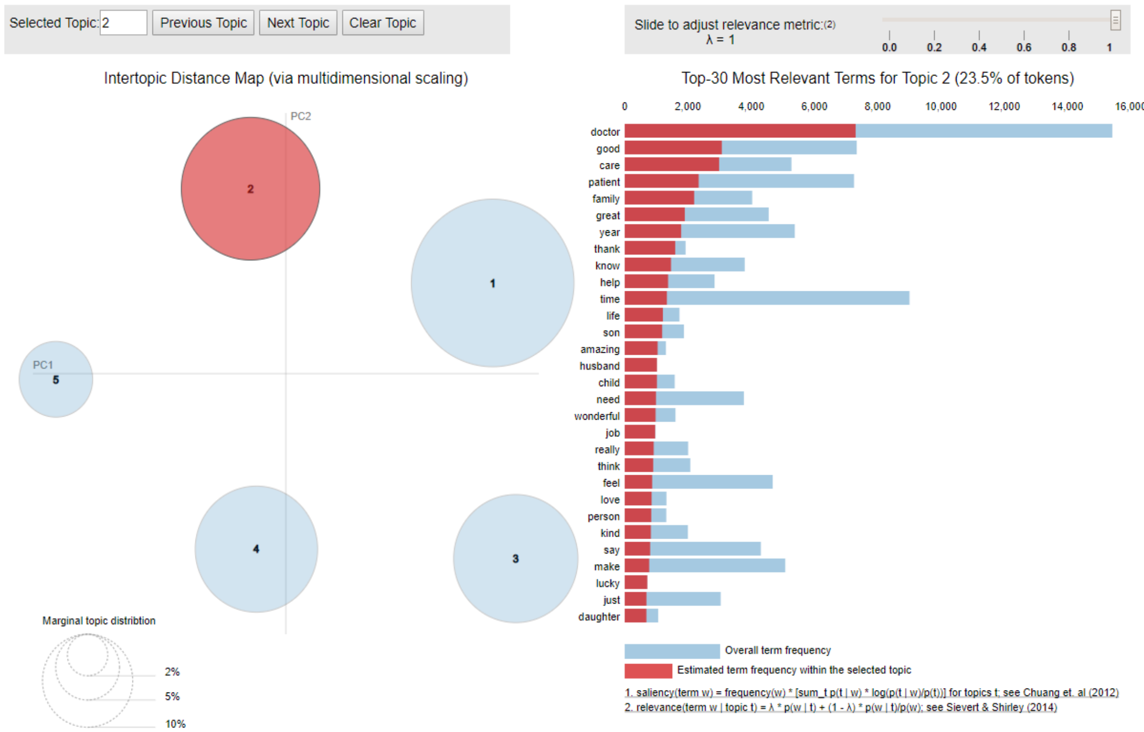

Figure 30: Number of Topics for 'Doctors in New Brunswick (NB)'.

Figure 31 depicts the learning decay for Doctors in New Brunswick (NB). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.7 and 0.9 outperform 0.5.

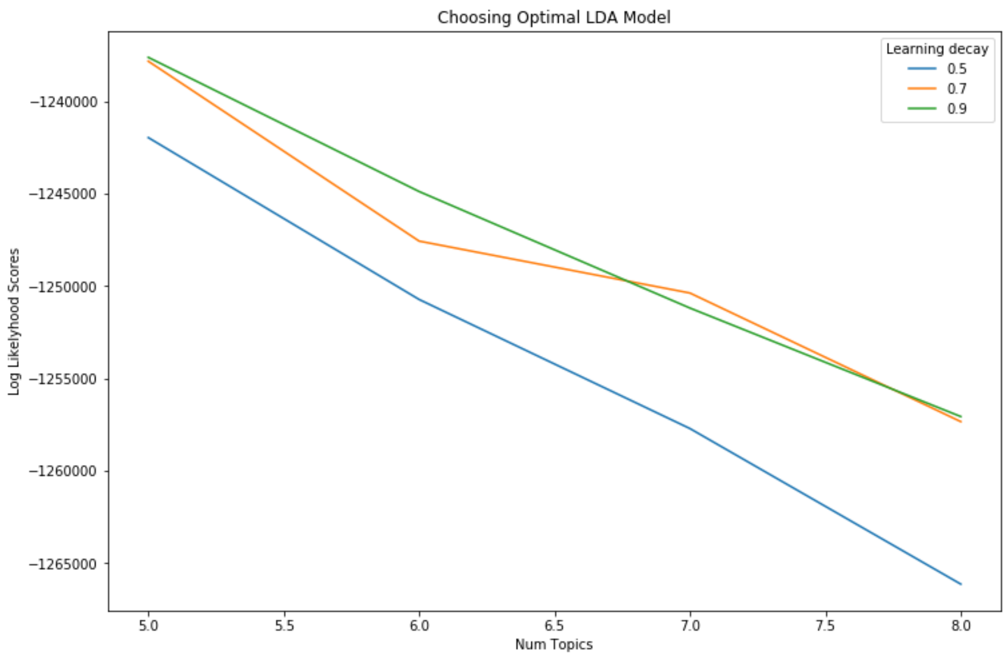

Figure 31: Learning Decay for 'Doctors in New Brunswick (NB)'.

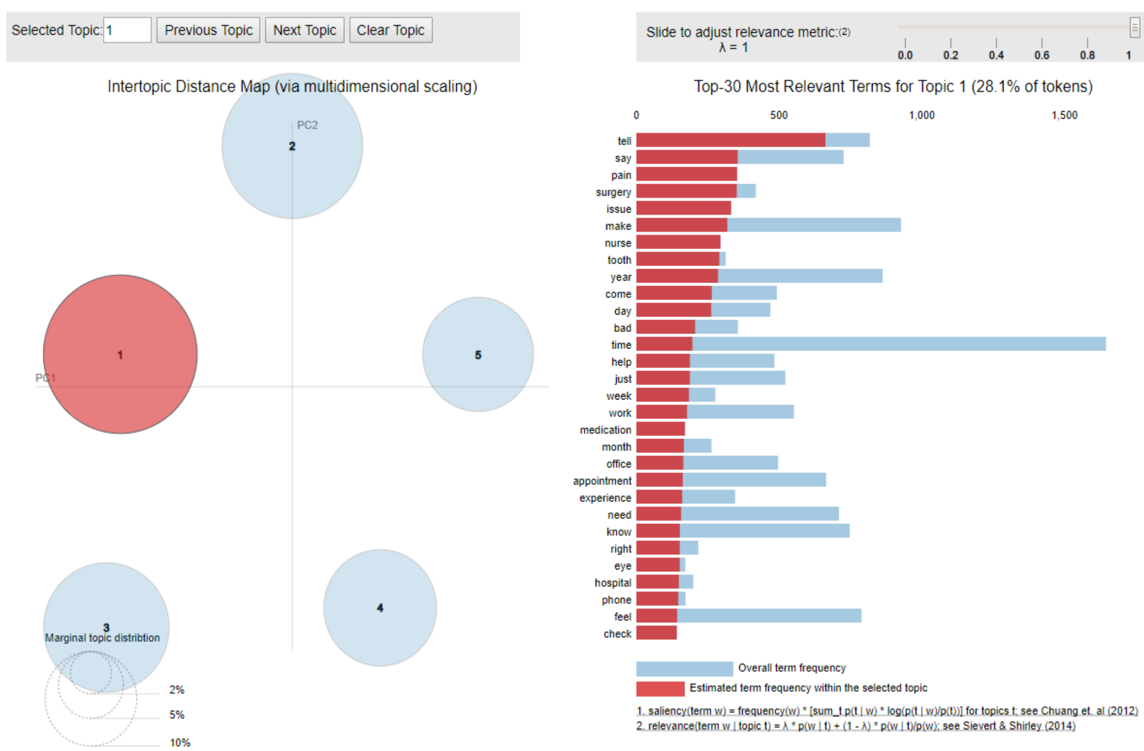

Figure 32: Number of Topics for 'Doctors in Prince Edward Island (PEI)'.

Figure 33 depicts the learning decay for Doctors in Prince Edward (PE). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

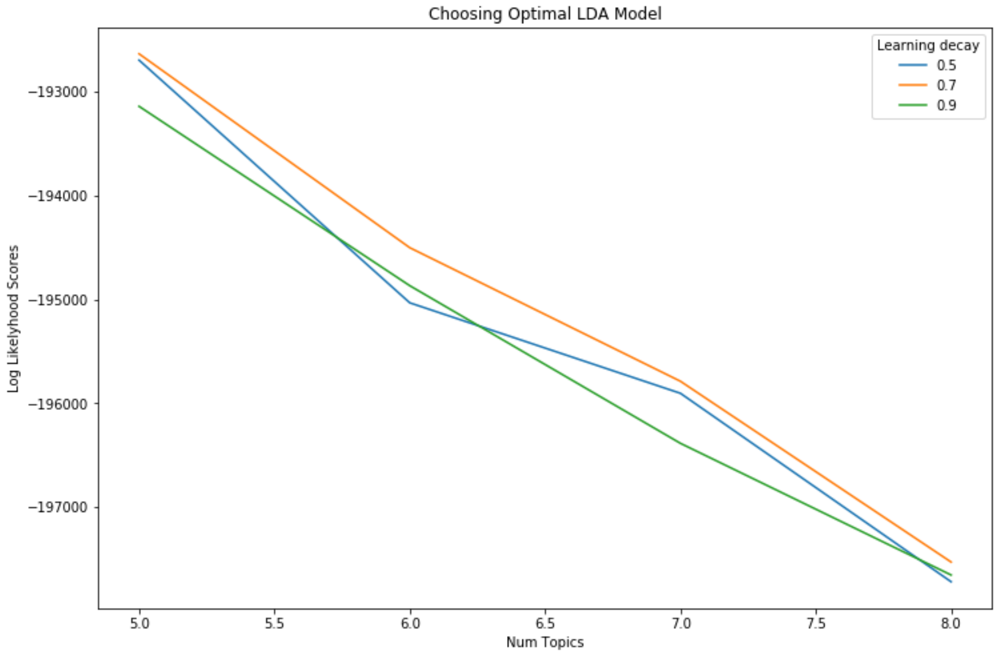

Figure 33: Learning Decay for ‘Doctors in Prince Edward Island (PEI)’.

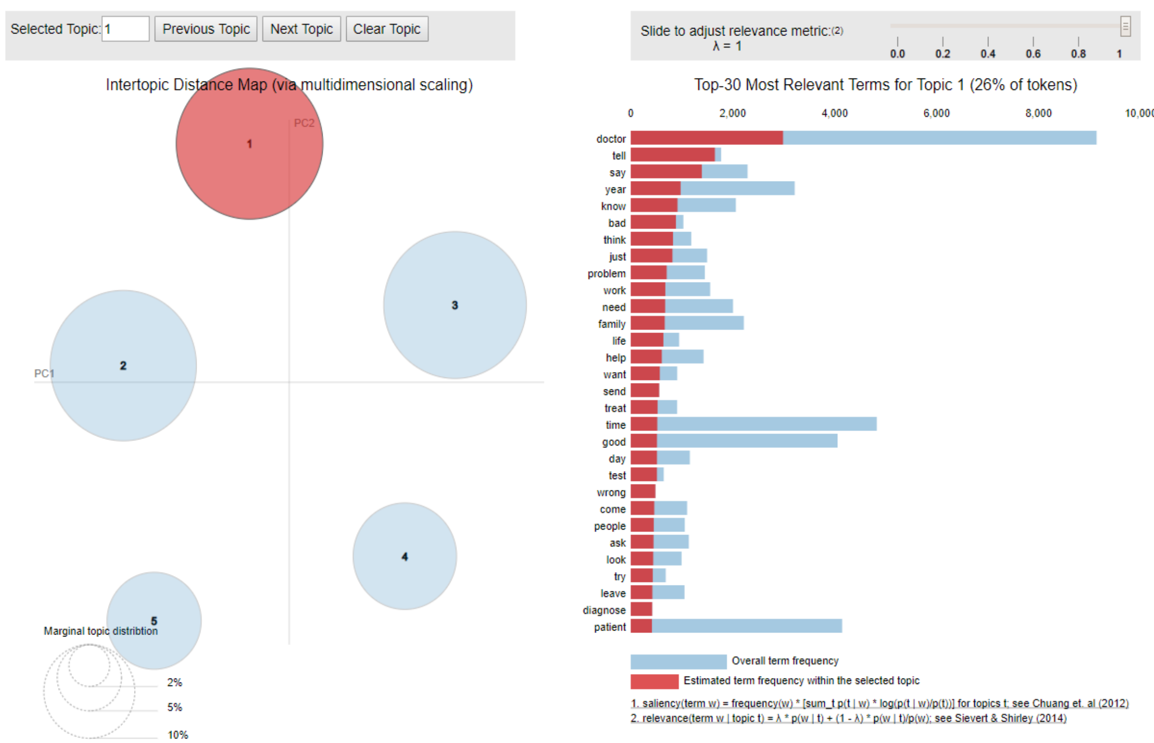

Figure 34: Number of Topics for ‘Doctors in Newfoundland&Labrador (NL)’.

Figure 35 depicts the learning decay for Doctors in Newfoundland&Labrador (NL). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.5 and 0.9.

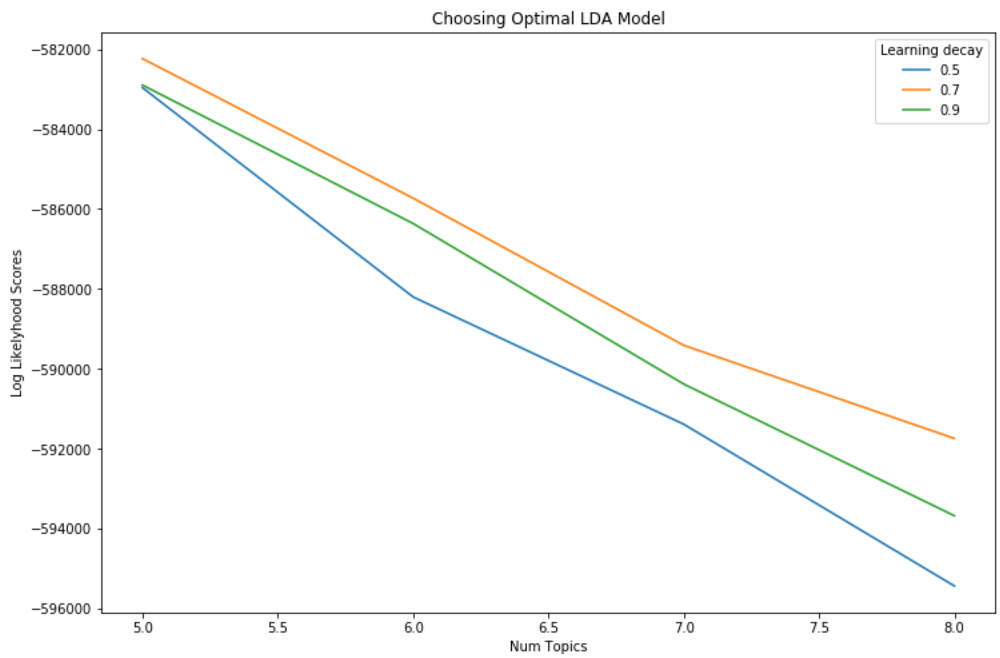

Figure 35: Learning Decay for ‘Doctors in Newfoundland&Labrador (NL)’.

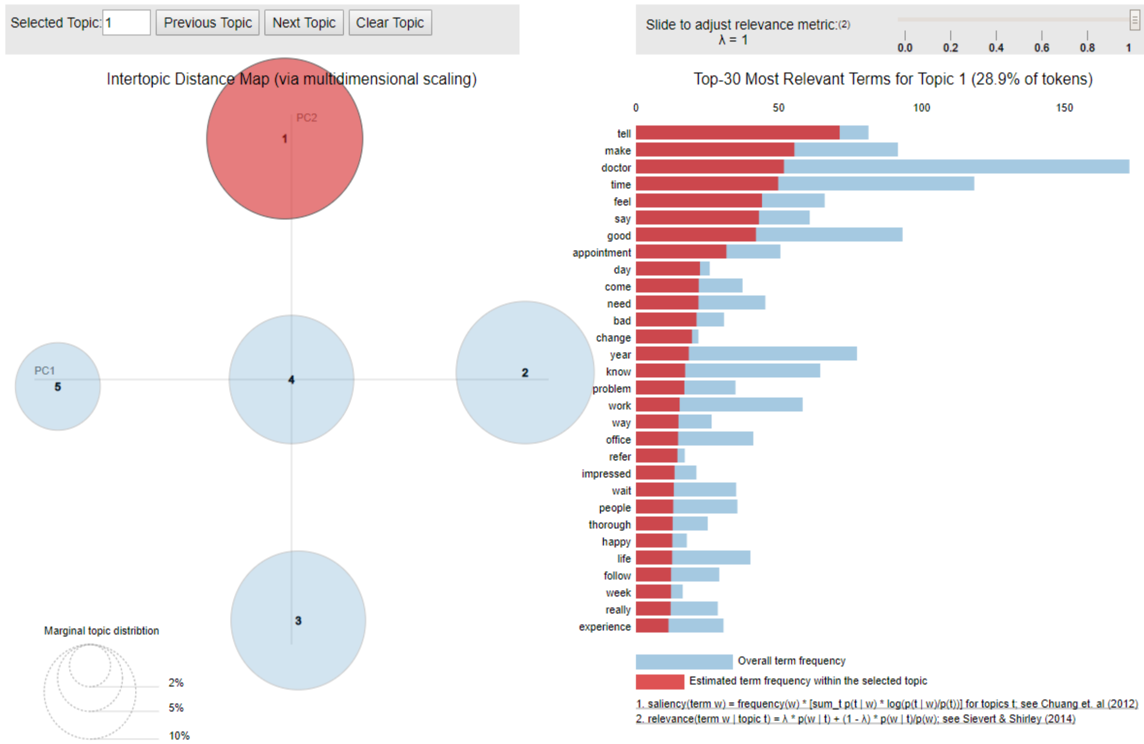

Figure 36: Number of Topics for 'Doctors in Saskechewan (SASK)'.

Figure 37 depicts the learning decay for Doctors in Saskechewan (SASK). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.9 outperforms both 0.5 and 0.7.

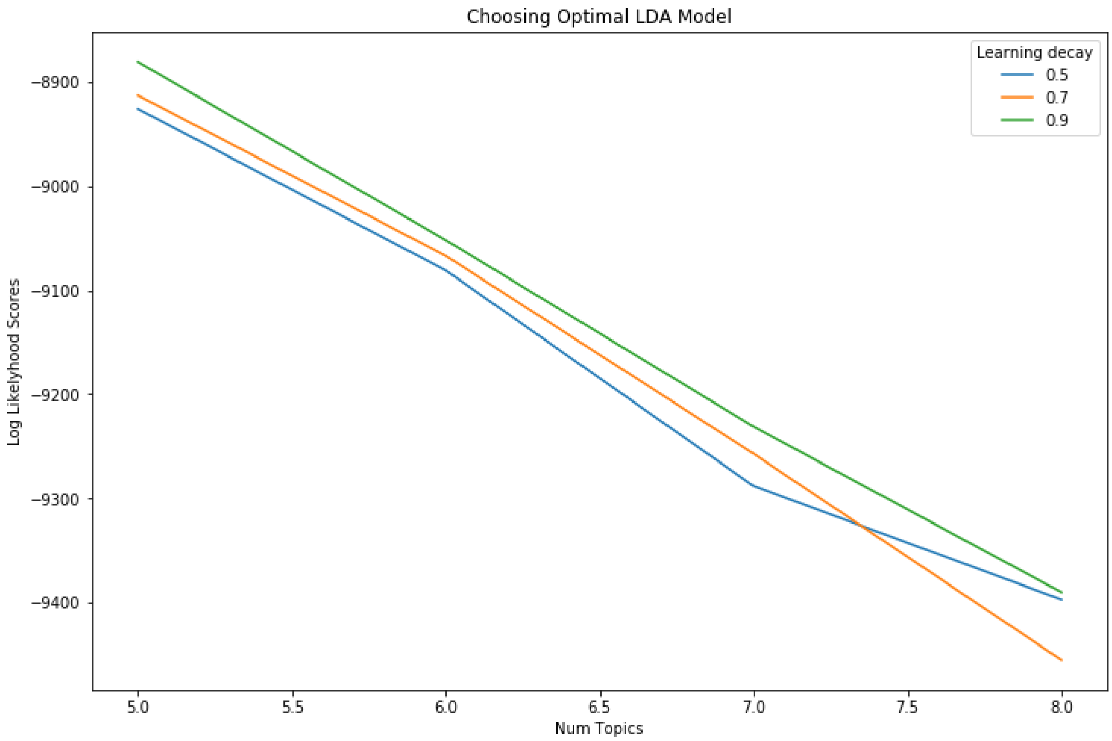

Figure 37: Learning Decay for 'Doctors in Saskechewan (SASK)'.

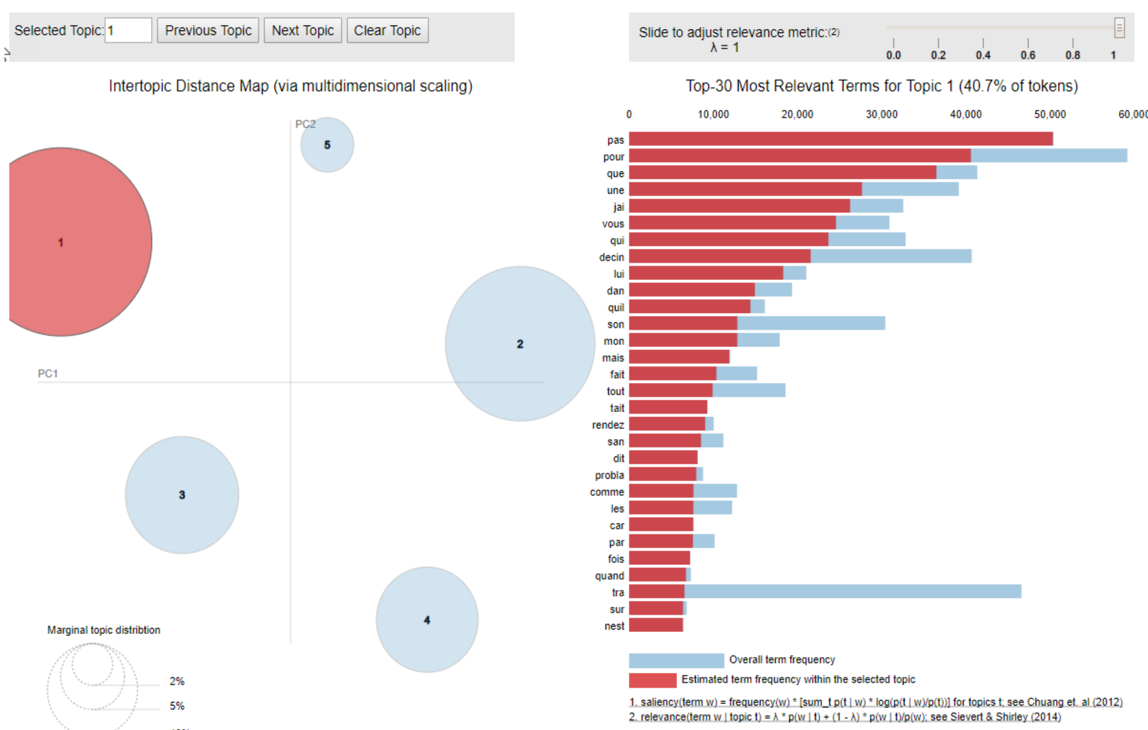

Figure 38: Number of Topics for 'Doctors in Quebec (QC)'.

Figure 39 depicts the learning decay for Doctors in Quebec (QC). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.7 outperforms both 0.7 and 0.9.

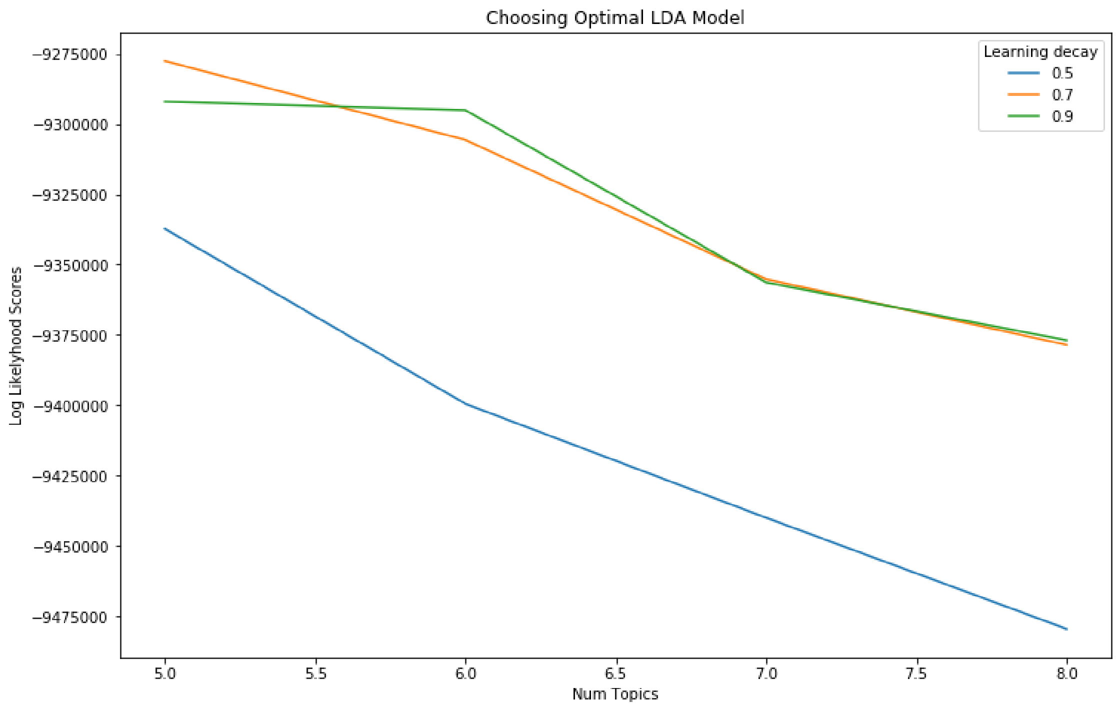

Figure 39: Learning Decay for 'Doctors in Quebec (QC)'.

Table 11: A breakdown of the territories, with the number of doctors, and number of text-reviews.

Territories	Number of Doctors	Number of Text-Reviews
Northwest Territories (NT)	68	670
Yukon	74	749
Nunavut (NU)	14	74

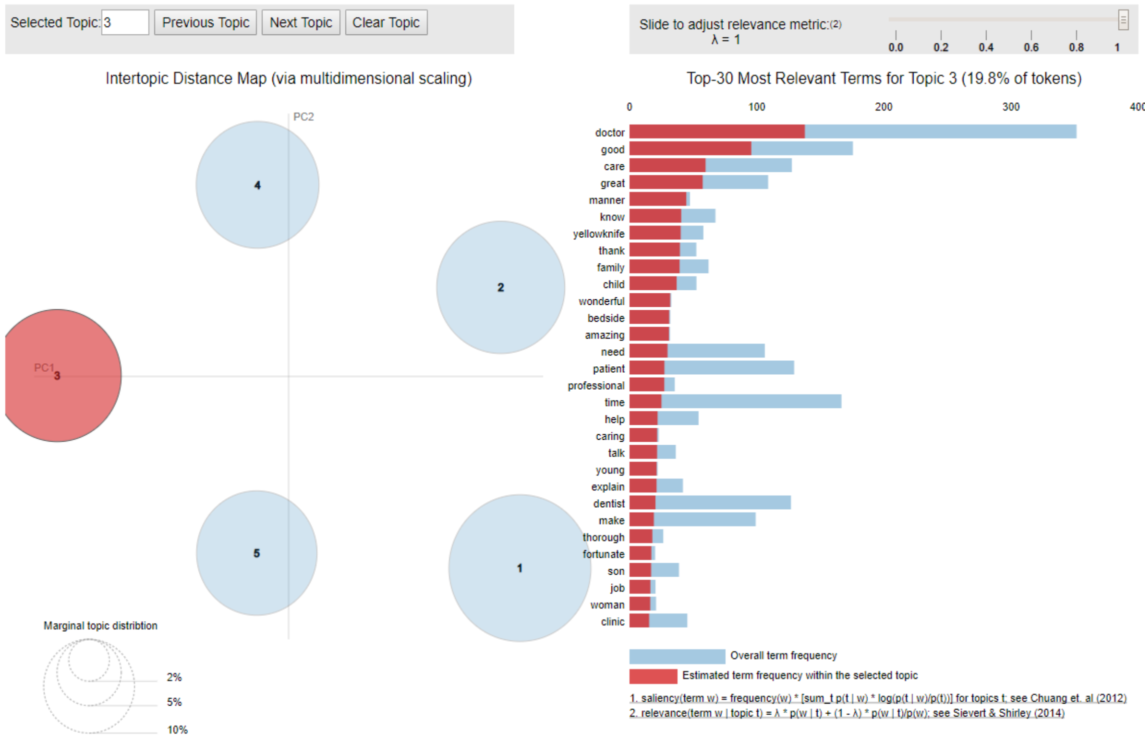

Figure 40: Number of Topics for 'Doctors in Northwest Territories (NT)'.

Figure 41 depicts the learning decay for Doctors in Northwest Territories (NT). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.5 outperforms both 0.7 and 0.9.

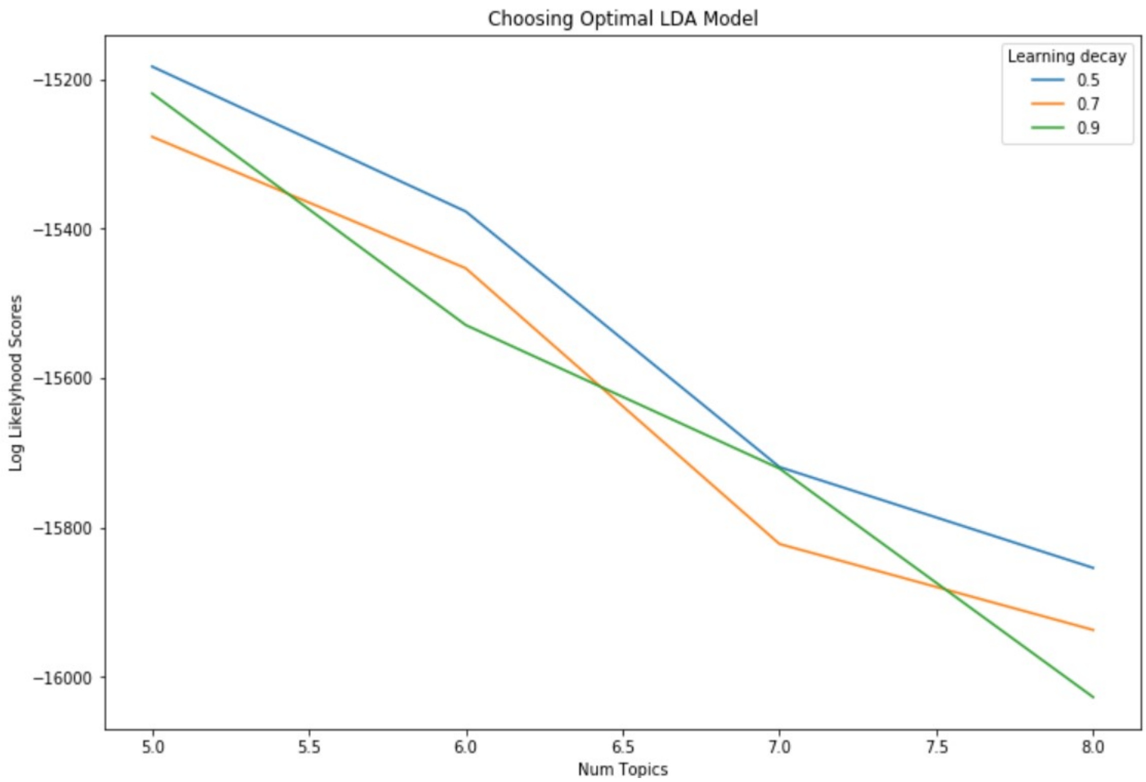

Figure 41: Learning Decay for ‘Doctors in Northwest Territories (NT)’.

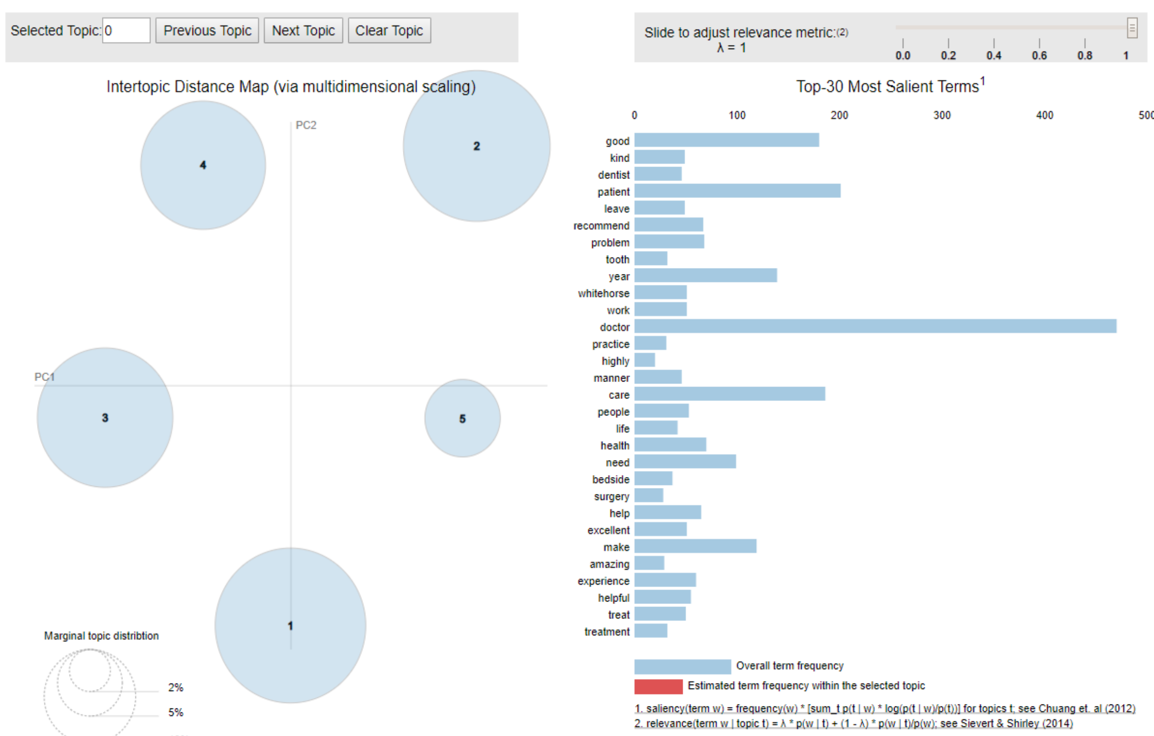

Figure 42: Number of Topics for ‘Doctors in Yukon (YT)’.

Figure 43 depicts the learning decay for Doctors in Yukon (YT). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.5 and 0.7 outperforms both 0.9.

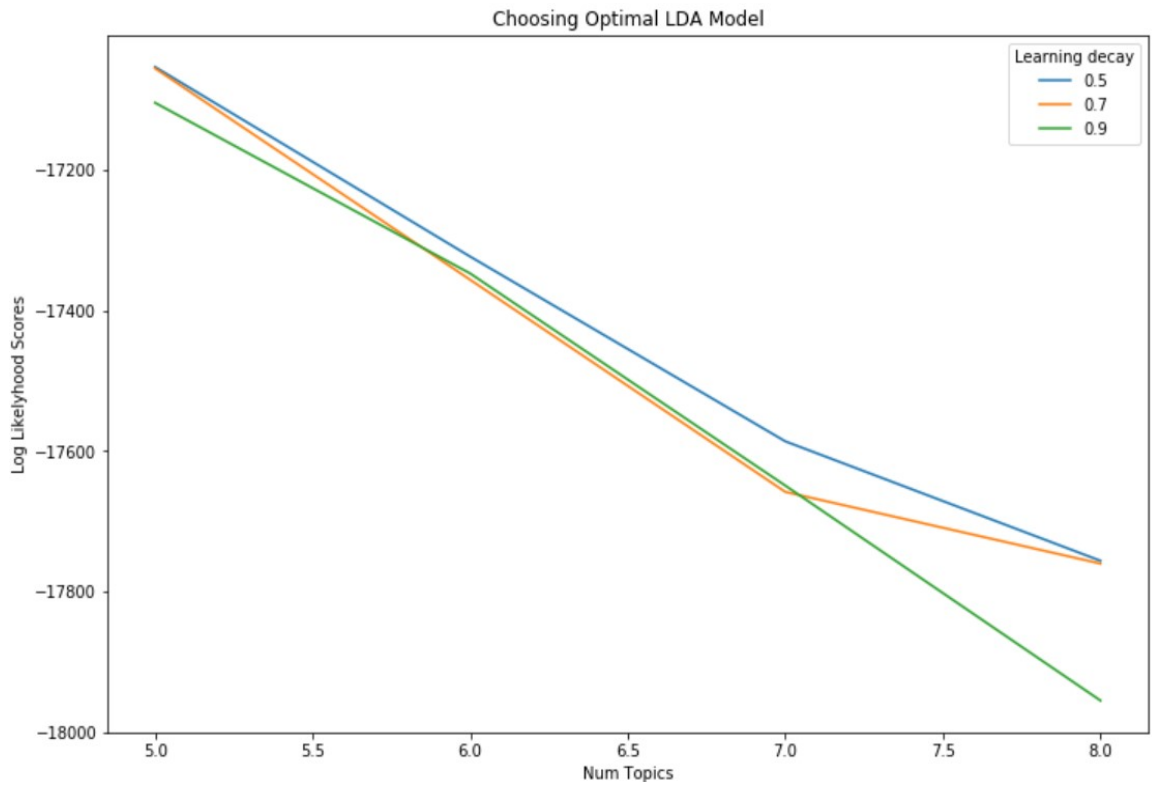

Figure 43: Learning Decay for ‘Doctors in Yukon (YT)’.

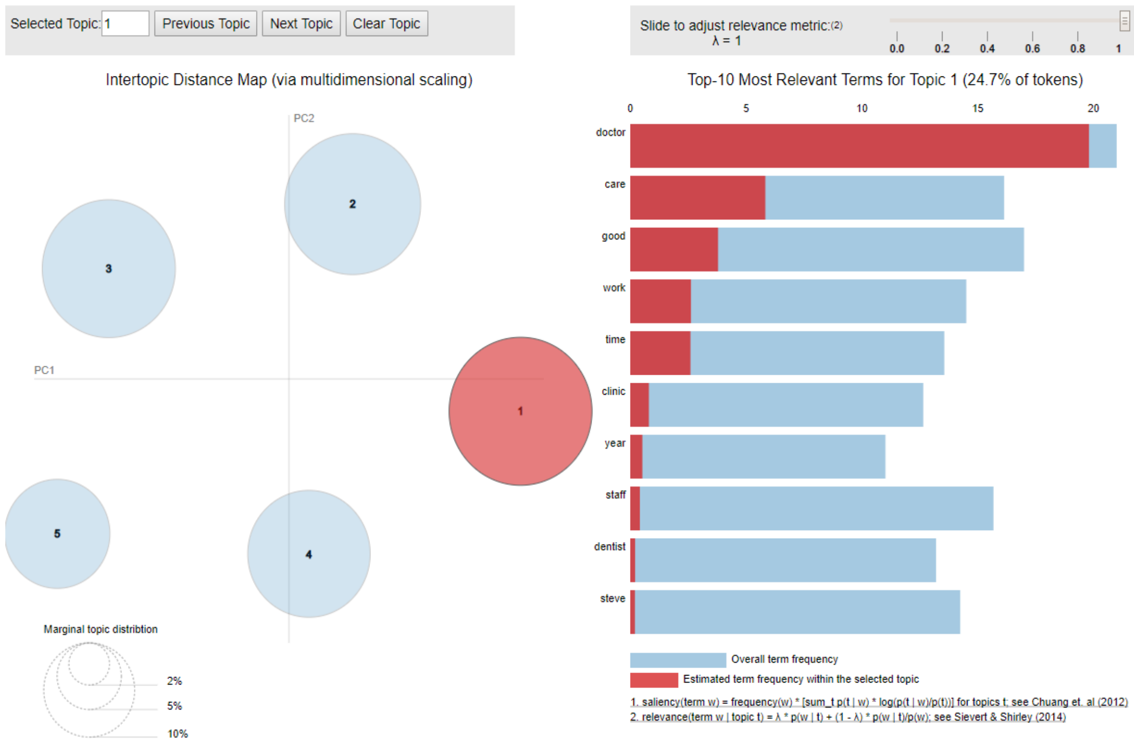

Figure 44: Number of Topics for 'Doctors in Nunavut (NU)'.

Figure 45 depicts the learning decay for Doctors in Nunavut (NU). As the output plot the log-likelyhood scores against num_topics, the plot clearly shows number of topics i.e. 5, 6, 7, 8. Furthermore, topic 5 has a better score. In addition, the `learning_decay` of 0.5 outperforms both 0.7 and 0.9.

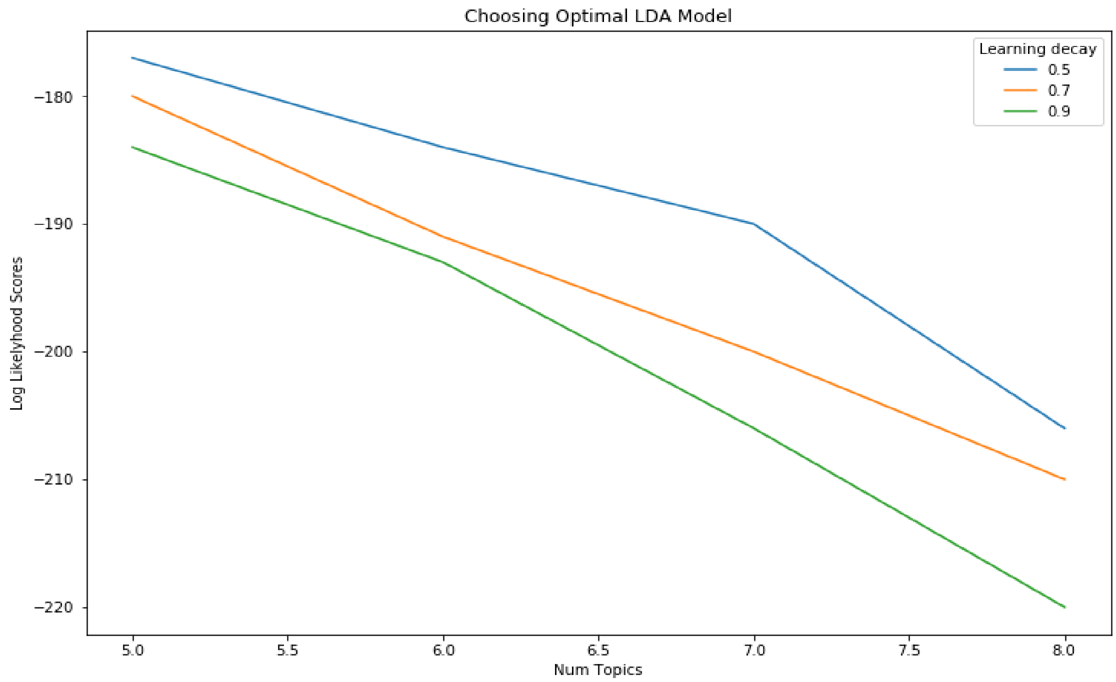

Figure 45: Learning Decay for ‘Doctors in Nunavut (NU)’.