**Analysing the social-ECONOMIC IMPACT of Wireless Mobile SERVICES DURING and before COVID-19 using topic modelling and Sentiment analysis on Tweets**

by

**[Wehel Hadi]**

A thesis submitted in partial fulfillment of the requirements for the degree of

**[Applied Data Science]**

**[University Utrecht]**

2021

**Approved by:** Dr. M.(Michael) Behrisch

Chairperson of Supervisory Committee

**Thesis supervisor:** Dr.A.A.A.(Hakim) Qahtan

**Second examiner:** Dr.M.W.(Mel) Chekol

**Program   
to Offer Degree:** Master’s degree

**Date:** 02/08/2021



**Table of Contents**

1. **Introduction**

Over the past twenty years, wireless communication has undergone a technological revolution. Wireless mobile services have become the fastest growing part of the telecommunications sector (Ronald Beaubrun, 2010). The use of a mobile phone has become an essential part of today's society. Today, 90 percent of the world's population over the age of six has access to a mobile phone, i.e. billions of people(Emerce, 2014). This makes wireless mobile communication a worldwide phenomenon, for developed countries as undeveloped countries. The rise of mobile technology therefore has a direct social-economic impact. The study of Subramani Parasuraman (2017) showed that a significant number of the participants of the research had an addiction to mobile phone usage. Another interesting result was that the majority of these participants didn’t recognized that they were addicted.

In contrast, wireless mobile services also have a positive social economic impact. Wireless mobile services, make a positive contribution to personal security. Nowadays it is possible to immediately ask for help in dangerous situations with the help of always available internet and mobile phone calls. In addition, the rise of this technology has also created a completely new industry. This has created many jobs for people, which has had a positive impact on employment and the economy (Ronald Beaubrun, 2010). This emerging industry is made up of different companies with different perspectives and ways of providing services. Two companies that differ in their way of delivering services are Infinity mobile and Mint Mobile. While Infinity mobile opts for the traditional fixed plan approach (Mobile I. , sd), Mint Mobile opts for more flexible plans for their customers (Mobile, sd).

THERE NEEDS TO BE MORE HERE

Therefore, this paper examines the social-economic impact of wireless mobile services for the companies Infinity Mobile and Mint Mobile. These two companies contradict each other in their way of delivering services. This paper will focus on the impact of wireless mobile services on user satisfaction (social effect), affordability (economic effect) and willingness to use (social effect). The analysis will be performed using topic modelling and sentiment analysis on data from twitter. Topic modelling will reflect the various topics discussed in the data. The purpose of this topic modelling analysis is determining the tweets that are related to the topics: user satisfaction, affordability and willingness. The sentiment analysis will determine what the sentiment is for each of these values. During this analysis, the changes in the sentiment score over time will be determined. Based on that, it will be clear whether COVID-19 affected the sentiment score or not.

The remainder of this paper is organized as follows: section 2 is the theoretical framework, where the two different companies are discussed and a brief introduction of the methods that are used for the analysis. Section 3 is the methodology, in this part the data collection, data cleaning, topic modelling and the sentiment analysis is discussed. Section 4 discusses the results, one part for the topic modelling and another part for the sentiment analysis. Section 5 contains the conclusion and section 6 contains the discussion. Section 7 describes the limitations of the research and the further research that needs to be done. The last section is the appendix.

1. Theoretical Framework

---------------------------------------------------------------------------------------------

* 1. Topic modelling

A topic model in machine learning and NLP is a statistical model that is able to distinguish the various topics that appear in the set of documents. This machine learning technique is able to automatically analyze text data and assign this data to a cluster (Pascual, 2019). Topic modeling is therefore an unsupervised machine learning technique; it does not require trained data that has already been classified in advance (Javatpoint, sd). Topic modeling is therefore a popular technique that is mainly used in natural language processing, this method can reveal the hidden semantic meanings of the data. As mentioned before, this technique will divide the data into clusters with the same words, these clusters are called the topics (Pascual, 2019). Topic modeling can be performed using various techniques. In this paper, the following three techniques will be applied: Latent Semantic Analysis/Indexing(LSA/LSI), Latent Dirichlet Allocation(LDA) and Hierarchical Dirichlet Process(HDP). These techniques will also be discussed below.

* + 1. Latent Semantic Analysis/Indexing(LSA/LSI)

Latent semantic analysis(LSA) is a statistical model that is able to determine the semantic word similarity between text data. Latent semantic analysis is also called latent semantic indexing, this is due to the purpose of the method, which is indexing text. The aim of this method is to improve the effectiveness of the matching of the semantic value of words (FOLTZ, 1996). This is achieved through query semantic matching instead of direct word matching. This makes it possible to recognize synonyms and thus assign them to the same semantic meaning. The idea behind this method is that there is a latent structure in the pattern of words usage across documents. This method also assumes that statistical techniques can be used to approximate this latent structure. However, this takes a number of steps.

Before this technique is able to analyze the text, the LSA method creates a matrix with the occurrences of each word in each document. The next step is that LSA uses singular value decomposition (SVD). SVD splits the original matrix into three smaller matrices, which after multiplication are equal to the original matrix. This is called the decomposition step. After this step, the three matrices are again further reduced in size, this is accomplished by choosing a smaller number of dimensions (Anandarajan, 2018). This process is shown below in figure 1.

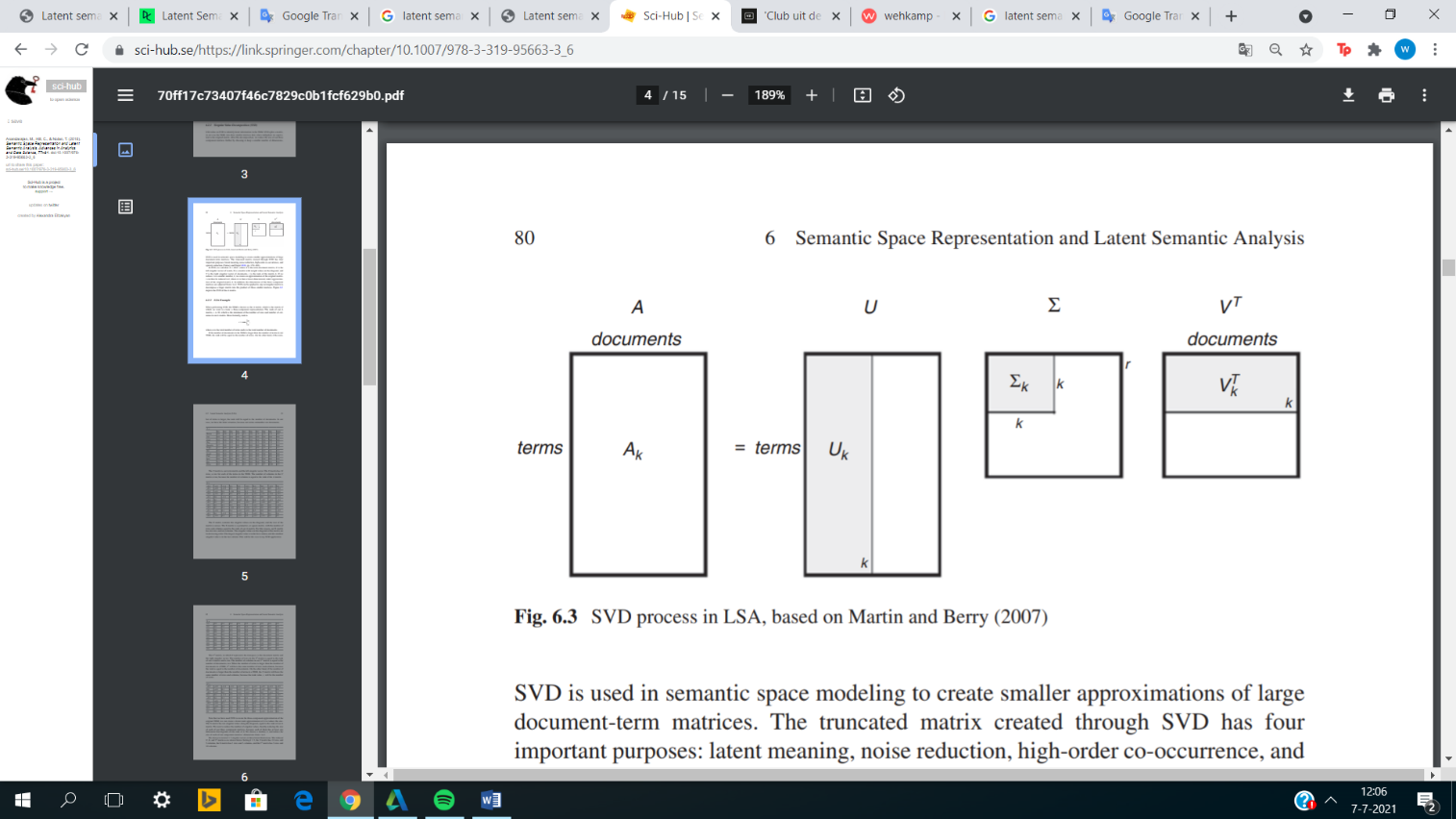


Figure 1. SVD process in LSA. (Anandarajan, 2018)

In the figure stands A for the original matrix, also called the term -document matrix. U stands for the left singular vector of words while V stands for right singular vector of documents and Σ is a weight value matrix. This gives the following formula:

* 1. .

R is the rank of the matrix; this method will approximate the original matrix during the dimension reduction (from r to k) (Anandarajan, 2018). A big advantage of this method is that for documents that don’t share any words in common, semantic matching is still possible.

* + 1. Latent Dirichlet Allocation(LDA)

Latent Dirichlet Allocation(LDA) is a probabilistic mixture model, which is used to find latent topics in text data. The word Latent in the name LDA shows that this model tries to find hidden topics from the documents. While the word Dirichlet indicates that the topics in the documents and the words in the topics have a Dirichlet distribution. Allocation refers to the distribution of topics in a document (David M. Blei, 2003).

This model assumes that each document contains a number of different topics and that the words in these documents are generated by these topics. Another assumption of this model is that each document has a different topic distribution (Team, 2020).

The idea behind this model is that the words in the documents help determine the different topics that appear in the document. Each word in the document is assigned to its topic. This allocation is determined using conditional probability estimates. This process is shown in the figure below.

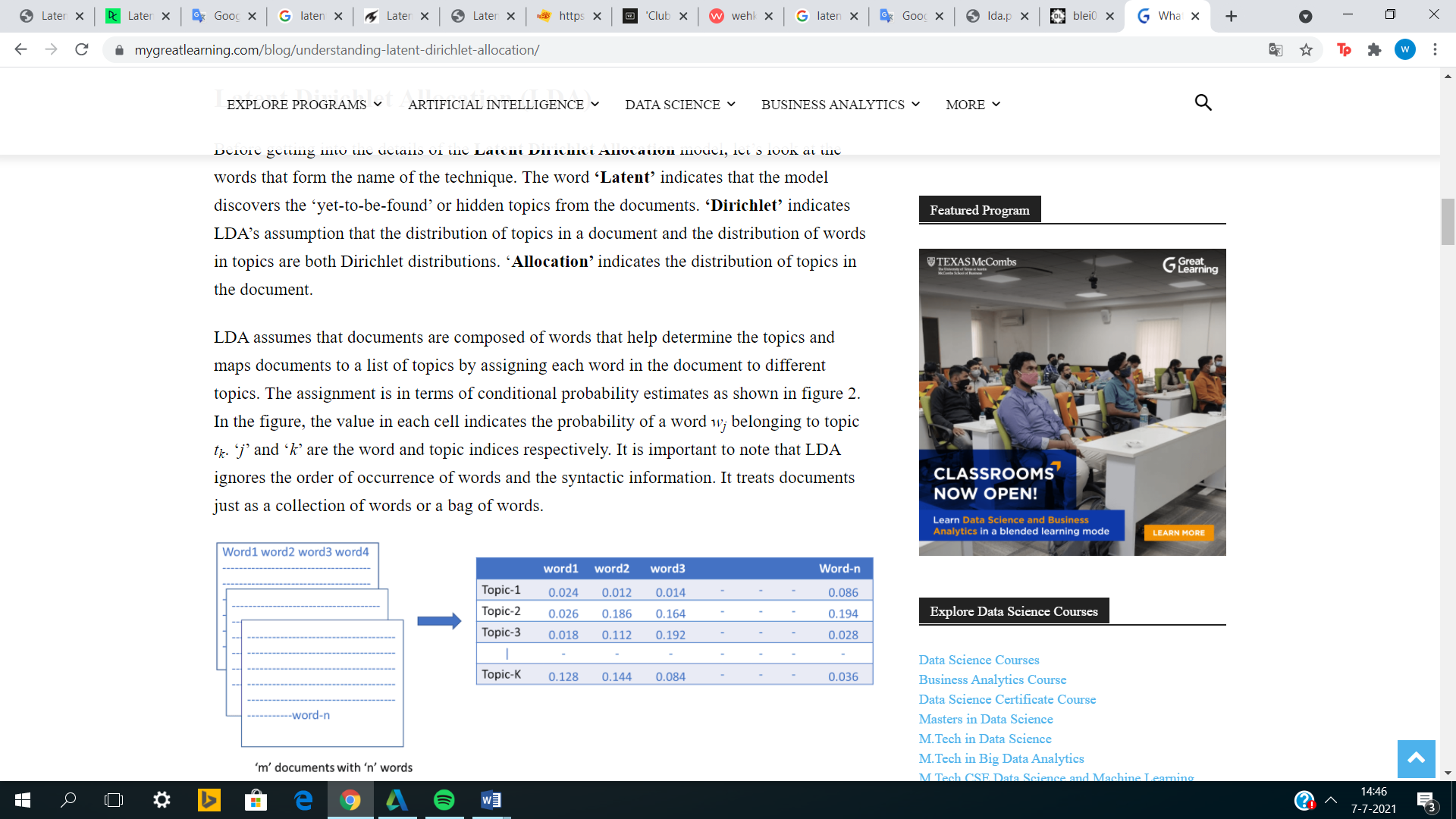


Figure 2. Topic probabilities for each word. (Team, 2020)

After the probabilities per word are known, the words must be assigned to the different topics. This can be done in two different ways. The first way is to set a certain probability threshold. As a result, only words with probability equal to or greater than this threshold will be assigned to the corresponding threshold. The other way is to say the top x probabilities of words are assigned to the corresponding topic (Team, 2020).

Determining how much a document belongs to a topic is done by the following formula:

* =

Determining how much a topic belongs to a word is done by the following formula:

* + 1. Hierarchical Dirichlet Process(HDP)

Hierarchical Dirichlet process(HDP) is the last topic model that has been applied. HDP is a model that clusters grouped data using a Bayesian approach. Like LDA, HDP uses a Dirichlet process for each group of data, meaning that the entire data has the same distribution, which is the dirichlet distribution. An advantage of this method is the statistical strength that arises because clusters contain data that belong to several groups of data. HDP shows many similarities with the LDA method, HDP is after all an extension of the LDA method. However, HDP has the great advantage that the number of topics do not have to be clear in advance. The disadvantage of this method is that it is difficult to apply, especially for projects where the number of topics do not necessarily have to be unbounded (Jordan, 2006).

* + 1. Evaluation of topic models

The above topic models should be compared. However, the question is: how to evaluate topic models? Topic models can be evaluated in three different ways.

- The first way is using human judgement. This can be observation based, namely by looking at the top N words in a topic. However, it can also be interpretation based, namely by looking at the words that do not belong to a topic, this is also called topic intrusion (Giri, 2021).

- The second way is using the following quantitative metrics: perplexity or coherence calculations. Perplexity is a measure for comparing probabilistic models. Perplexity is an indication of how well a probabilistic model is able to predict a sample. Usually, the lower this perplexity value is (around 0) the better the topic model functions. Coherence is a measure for the semantic similarity between words in topics. The coherence score varies between 0 and 1, but generally the higher the coherence score, the better the topic model. The coherence scores and the respective valuation are shown below (Giri, 2021).

|  |  |
| --- | --- |
| Coherence scores | Valuation |
| 0.3 | Bad |
| 0.4 | Mediocre |
| 0.5 | Sufficient |
| 0.65 | Great |
| 0.65> | Unrealistic |

These coherence scores are determined using the following formula (Nikolaos Aletras, 2013):



- The 3rd way is the combination of human judgment and quantitative metrics.

* 1. Sentiment analysis
     1. Logistic Regression classifier
     2. Naïve Bayes classifier(TextBlob)
     3. Decision trees classifier
     4. Evaluation of sentiment models

1. Theoretical Literature

Wireless mobile services are indispensable in our society today, but this has not always been the case. The wireless technology industry has evolved over the years, leading to a mobile revolution with fundamental and major consequences for the world. However, this revolution has not yet come to an end, innovations are made every year. This requires adaptability from both society and the economy. However, the advantages of these emerging technologies outweigh the disadvantages. This revolution has made it possible to obtain more easily available data so that decision makers can make better decisions. It has also had a positive influence on communication between people, but also between institutions and people. The government has been given a tool that makes it easier to reach the people. Companies also have this advantage with regard to their customers (Keng Siau, 2003). The development of this sector has gone through several intermediate steps. These step will be discussed in chapter 3.1.

* 1. Evolution of wireless mobile services

The wireless mobile services revolution started in 1980, since then this technology has been developing constantly and rapidly (Keng Siau, 2003). This mobile technology has gone from 1G in 1980 to 5G in 2020. This is shown in figure 3.

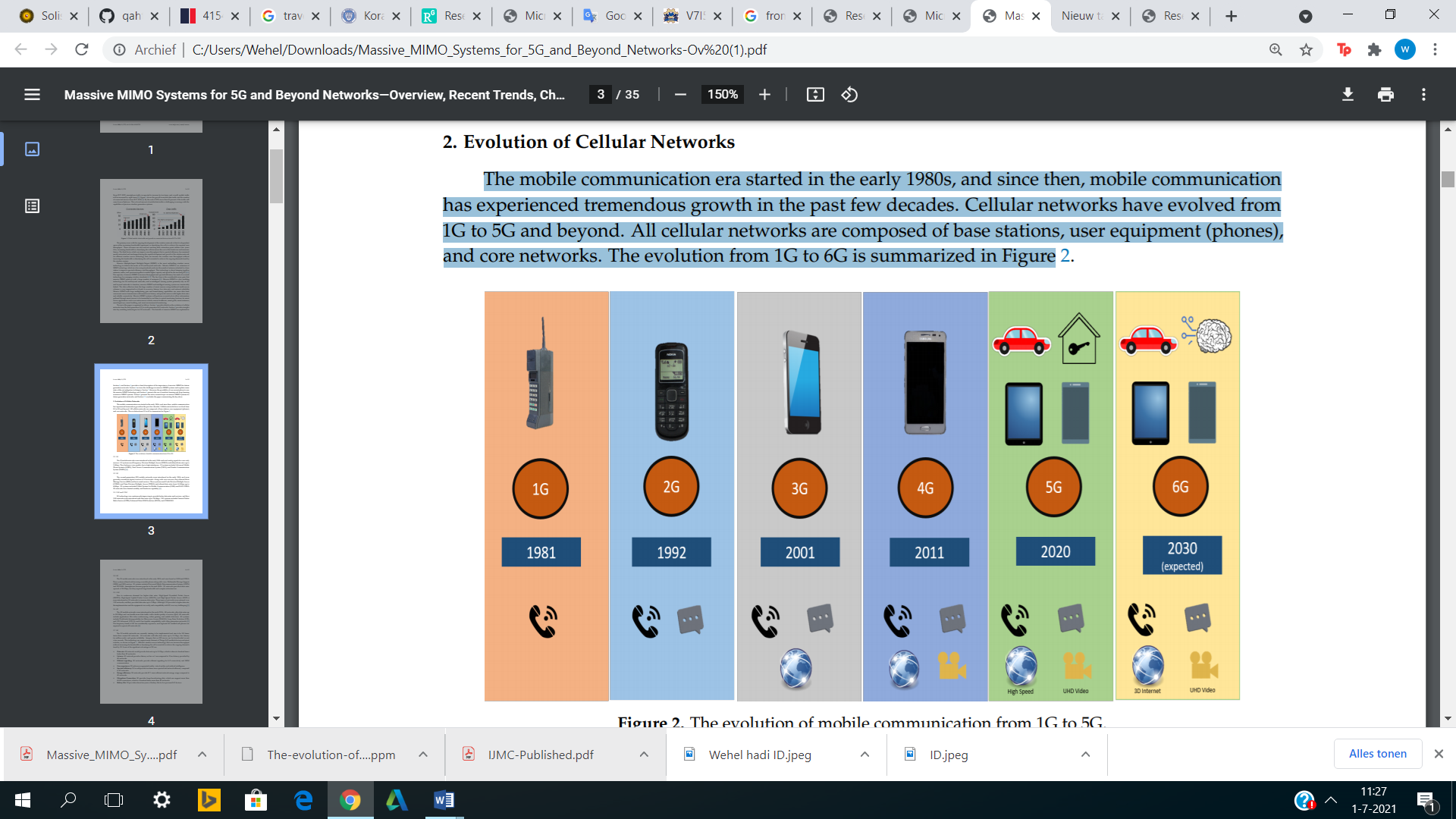
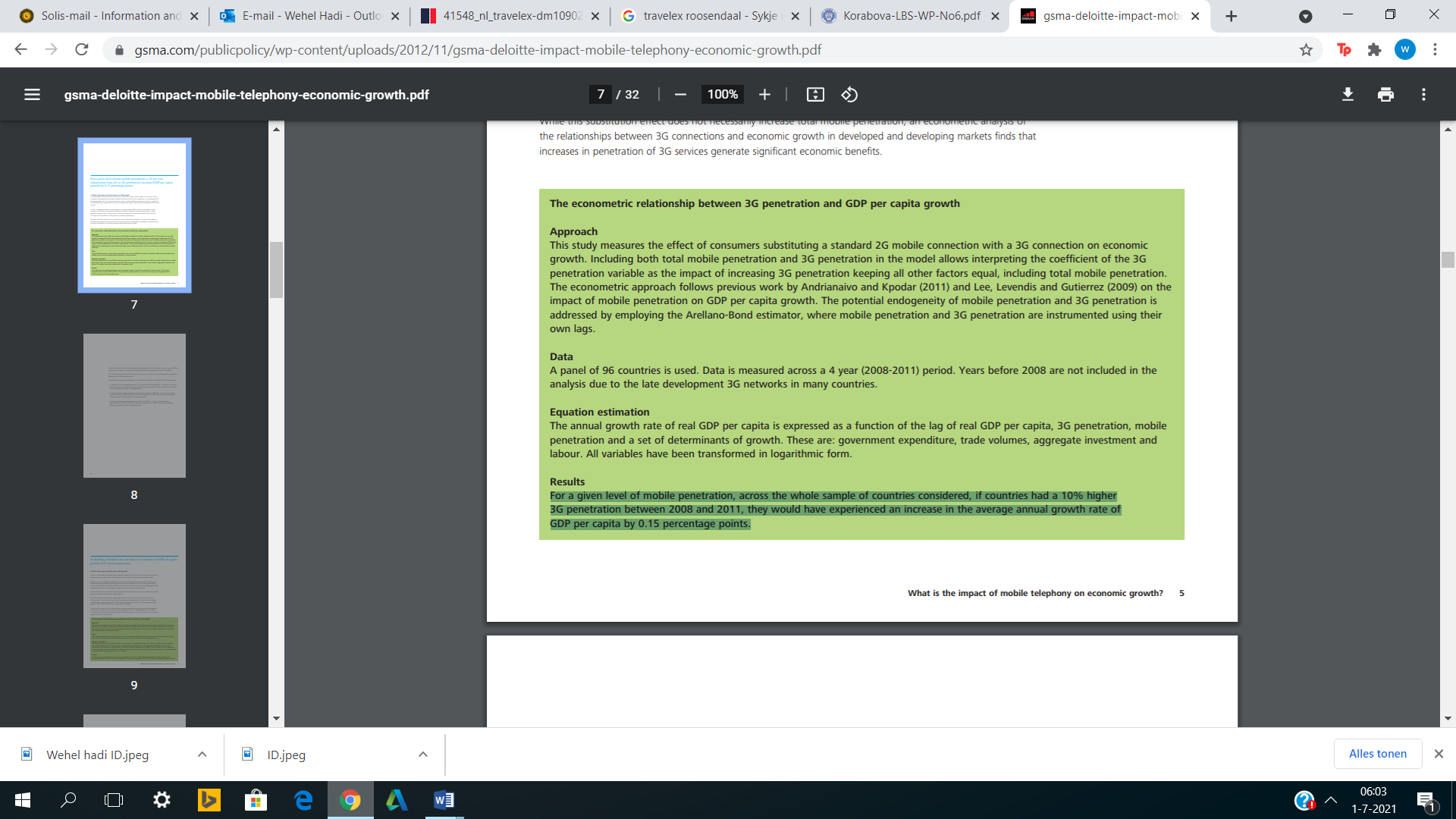


Figure 3: The evolution of mobile communication from 1G to 5G (Chataut, 2020)

However, the basic concept for each of these innovations is the same, namely that the cellular network consists of a base station, user equipment(phone) and core networks (Chataut, 2020). The enormous demand by consumers, companies and governments for these mobile services has accelerated the developments around mobile technology. The various phases that led to this development are described below.

* In the beginning of 1980 the 1G mobile networks came on the market. This network used analog signals for voice-only services. This network also provided data rates up to 2.4 kbps. This was the beginning of this technology and because of this there were many shortcomings in the wireless mobile service. A big disadvantage of these 1G mobile networks was that it had a poor voice quality, this was due to high interference (Chataut, 2020).
* At the beginning of 1990, the 2G mobile networks came on the market. The 2G networks were more like a digital version of the 1G networks from the 1980s. This made it possible to send a text message and a number of simple email services with a mobile in addition to the voice services. This mobile network also provided data rates up to 64 kbps. Despite the fact that the 2G mobile networks were an improvement compared to the 1G mobile networks, there were still shortcomings. A major drawback of this network was that it had limited mobility and hardware capability (Chataut, 2020).
* In the early 2000s, the 3G mobile networks came on the market. The system behind this network was called GSM. This system made it possible to browse the web in addition to calling and texting. This network also had a data rate up to 384 kbps. This system was another major step forward, but the major drawback of this network was that it required a large bandwidth (Chataut, 2020).
* At the beginning of the 2010s, 4G mobile networks came on the market. This network has many similarities with the 3G network, but it is able to handle more data with a better service. The data rates of this network can reach up to 100 Mbps. In contrast to the older networks, this network makes it possible to play online games, make video calls and watch television on your mobile. However, this network has the disadvantage that the frequency bands are expensive and that customers need a high tech telephone to make this network work(which is expensive) (Chataut, 2020).
* The 5G
  1. Social-economic impact of wireless mobile services
  2. Mint Mobile
  3. Infinity Mobile



<https://stackoverflow.com/questions/10059594/a-simple-explanation-of-naive-bayes-classification>

1. Methodology
2. Results
3. Conclusion
4. Discussion
5. References