HEC MONTRÉAL

Computer aided literature review in the context of Fintech for financing SMEs

par

Quan Hao

Sciences de la gestion
(Option Thesis)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

December 2019

Abstract

In this thesis, we propose the use of text mining to achieve a computer aided literature review. By using simple keywords as “SME financing difficulties” and “Fintech SME solution”, a set of papers were obtained related to fintech and financing SMEs. We then propose to apply text mining in order to classify the so obtained papers and extract the most important information from those documents before reading them. The researcher may then focus on one aspect of the problem with no need to read all the document of select arbitrarily some of them.

Keywords

SMEs, Financing, Text mining

Research methods

Clustering, K means, Hierarchical clustering, SVD, TF-IDF

Contents

List of Figures

Introduction

The goal of the thesis is to summarize the financing difficulties of SMEs and the Fintech as a new tool to resolve the financing difficulties of SMEs based on the clustering algorithms. More precisely, we propose to use text mining to support researchers for building literature review.

To present our research, we divided the thesis into five chapters. The first chapter will focus on the context of SMEs/banks/Fintech. In this chapter, we will use the situation of SMEs in Canada as an example to illustrate the importance of SMEs in a country's economy as well as the current status of SMEs' financing. In addition to that, we will conclude the reasons that cause the current status of SME financing. We will also compare the differences between traditional and Fintech financing. The second chapter is the literature review on web mining and text mining. We present the main ideas of web mining and text mining in the context of clustering algorithms. The third chapter is the technical part of the thesis. To start our search, our keywords are set to "SME financing difficulties" and "Fintech SME solutions." We also analyze the structure of web pages and illustrate how we scrape the pdf files and HTML files as our dataset from the internet. After scraping the pdf files and HTML files, we explain how we proceed with the clustering algorithms. To conduct the clustering algorithms, we first built a corpus and then preprocessed our dataset. After getting the "clean" dataset, we select the best K based on the elbow curve. We used the best K on K means and hierarchical clustering to get the clustering results. The fourth chapter will present the description of each cluster and make an overview to links of each cluster. The fifth chapter is the conclusion of the thesis, and we will also present the ideas for future researchers.

Chapter 1

Context

1.1 Importance of SMEs

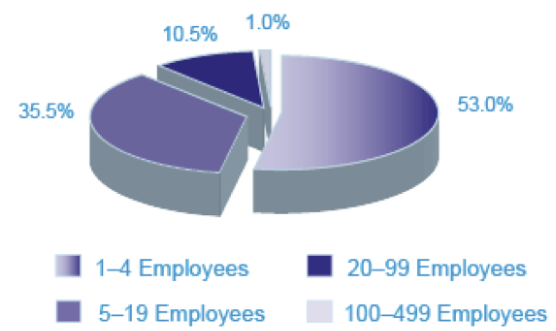
In the American economy, there are 30 million small businesses representing nearly 99 percent of all American companies (Mills and McCarthy 2016). In Canada, SMEs also play an important role in the economy, and the following will present the importance of SMEs in the Canadian economy.

According to criteria from the Canadian government (2013), a business establishment must meet one of the following criteria: have at least one paid employee (with payroll deductions remitted to the Canada Revenue Agency (CRA)), have annual sales revenues of $30,000, or be incorporated and have filed a federal corporate income tax return at least once in the previous three years.

The data (Industry Canada 2013) is based on the businesses that are neither self-employed nor "indeterminate." It also does not take industrial sectors into account. These industrial sectors include public administration (schools and hospitals), public utilities and non-profit associations. Although there are many ways to define the size of the enterprise, the data in the following context is defined based on the number of people employed by the company. According to the concept from Statistics Canada (2013), SME is a business establishment with 1-499 paid employees.

From the classification of enterprises, we can have a general idea of different sizes of enterprises from figure 1, and 99% of the entire business in Canada can be defined as a small business.

Figure 1: Distribution of SMEs by Size of Business, 2011



Source: Statistics Canada, *Survey on Financing and Growth of Small and Medium Enterprises*, 2011.

From different sectors' perspectives, we can get more detailed information about SMEs. According to Statistics Canada (2019), as of December 2017, there are a total of about 1.18 million businesses in Canada. Among these businesses, 97.7% are small enterprises, which are 1.15 million, and 1.9% are medium-sized businesses, which are 21926 medium-sized enterprises. About 0.2% are large enterprises. Figure 2 shows a detailed picture, and we can see that three out of four business has 1-9 employees. We can also know that in both of service industry and goods industry, the number of SMEs accounts for the vast majority of the total number of enterprises (Innovation, Science and Economic Development Canada 2019).

Although SMEs occupy the majority of the total businesses. It can be seen in figure 3 that SMEs tend to decline as the number of years increases. The horizontal axis of the graph indicates that the years of the establishment of the business on average based on the initial business size.

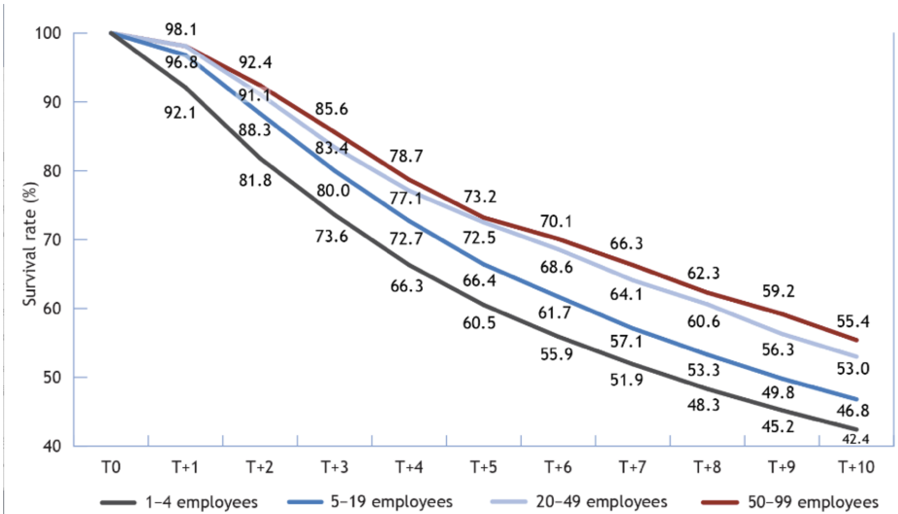
Figure 2: Number of employer businesses by sector and business size(number of employees), December 2017

Number of employees	Goods		Service		Total	
	Number	Cumulative %	Number	Cumulative %	Number	Cumulative %
1–4 employees	144,678	56.9	489,385	53.0	634,063	53.8
5–9 employees	49,059	76.2	181,798	72.7	230,857	73.4
10–19 employees	27,736	87.2	125,065	86.2	152,801	86.4
20–49 employees	19,723	94.9	81,630	95.1	101,353	95.0
50–99 employees	7,049	97.7	26,646	97.9	33,695	97.9
Small businesses *(1–99 employees)*	*248,245*	*97.7*	*904,524*	*97.9*	*1,152,769*	*97.9*
100–199 employees	3,526	99.1	11,316	99.2	14,842	99.1
200–499 employees	1,797	99.8	5,287	99.7	7,084	99.8
500+ employees	538	100.0	2,401	100.0	2,939	100.0
Total	**254,106**	**21.6**	**923,528**	**78.4**	**1,177,634**	

Note: By definition, the goods-producing sector consists of agriculture, forestry, fishing and hunting; mining, quarrying, and oil and gas extraction; utilities; construction and manufacturing. The service-producing sector consists of wholesale trade; retail trade; transportation and warehousing; information and cultural industries; finance and insurance; real estate and rental and leasing; professional, scientific and technical services; management of companies and enterprises; administrative and support, waste management and remediation services; educational services; health care and social assistance; arts, entertainment and recreation; accommodation and food services; other services (except public administration) and public administration.

Source: Statistics Canada, Table 33-10-0037-01 — Canadian Business Counts, with employees, December 2017.

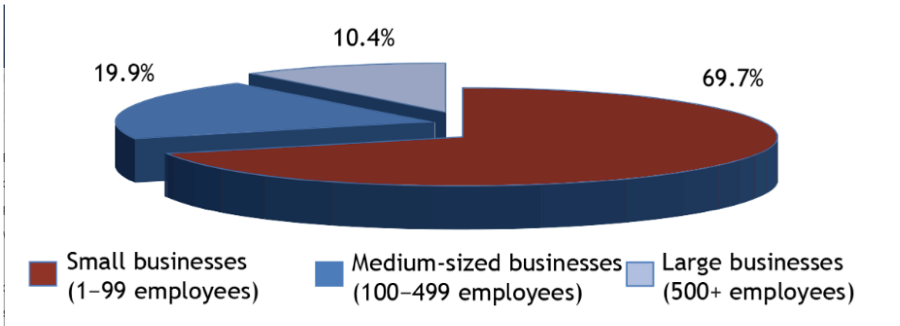
Figure 3: Survival rate by initial business size



Source: Statistics Canada, Canadian Centre for Data Development and Economic Research, National Accounts Longitudinal Microdata File.

From the number of workforce's perspective, it can be seen that SMEs also employ the majority of workforce. In 2017, private companies employed about 11.9 million people. Most of them work for small businesses. As it can be seen from figure 4, small businesses attract the largest number of workforce, which are approximately 69.7% (8.3 million) of the workforce, and about 19.9% (2.4 million) of the workforce work for medium-sized companies. Large companies only attracted about 10.4% (1.2 million people). There are about 89.6% (10.7 million) of the employed people working in small and medium-size companies (SMEs).

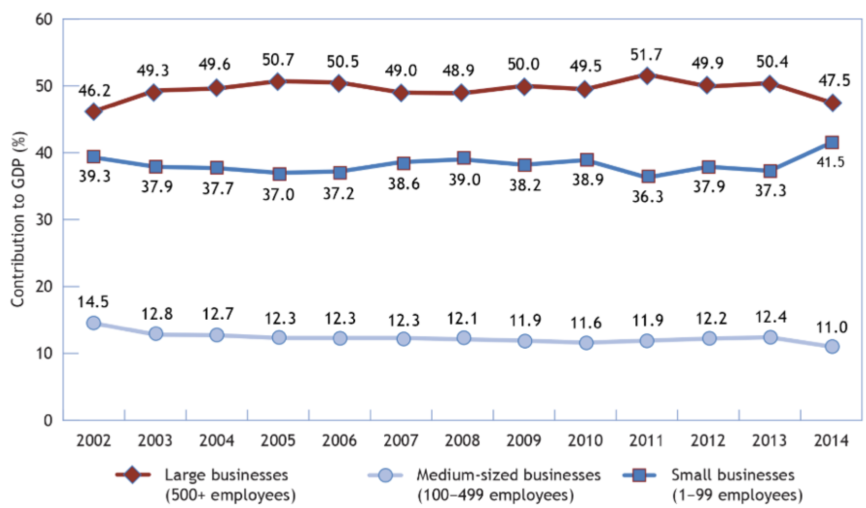
Figure 4: Distribution of private sector employees by business size, 2017



Source: Statistics Canada, *Labour Force Survey*; and ISED calculations.

SMEs not only play a vital role in the number of total businesses and the total employed population but also has an essential impact on the contribution rate of GDP. Statistics Canada (2019) has drawn a graph of the growth of GDP in the private sector from 2002 to 2014, as shown in the figure 5. From the overall trend, it can be seen that from 2002 to 2014, the ranking of the contribution rate of GDP in three different sized businesses has not changed much. From the average data from 2010 to 2014, it can be seen that the contribution rate of small enterprises to GDP is 38.4%. The contribution rate of medium-sized enterprises to GDP is 11.8%. Large enterprises that contribute the most to GDP accounted for 49.8%. Although large companies accounted for the most considerable contribution rate of GDP, SMEs exceeds the 50% contribution rate of GDP.

Figure 5: Contribution to GDP by business size, Canada, 2002-2014



Source: Statistics Canada.

1.2 Current Status of Small and Micro Enterprise Financing

Statistics Canada (2013) conducted a related survey on SME financing in 2011. The survey data was sampled by 10,000 survey respondents showing that about 36% of SMEs demanded external financing and obtained 36$ million of capital.

According to Statistics Canada (2013), all kinds of companies have financing needs. For example, in 2011, about 39% of start-up businesses have financing needs. There are also financing needs for businesses that have been established for more than 20 years, and these companies account for 34% of the total financing companies.

Start-up SMEs lack credit lines and therefore, the demand for loans from financial institutions is generally not met. Only about 40.9% of businesses received credit from financial institutions.

Overall, 36% of SMEs require external financing. Many startups choose not to seek external financing. According to the data, 6.6% of companies that have operated for two years or less have not chosen external financing, while only 1.2% of companies that have been in business for more than 20 years have chosen to finance.

We can see the basic situation of financing needs according to the number of years of enterprise establishment from figure 6. The horizontal axis represents the debt requested rate, and the vertical axis represents the authorized-to-requested. It can be seen from figure 6 that the startup that is established in two years or less is the hardest to get financed. There is a 27.5% request rate in this category, but there are only 81% of authorized-to-requested. For companies that have a more extended establishment period, e.g. establishment of 2 years to 10 years, The 27.1% request rate corresponds to 86.3% authorized-to-requested.

Figure 6: Authorized-to-Requested Ratio versus Debt Request Rate by Category of SME, 2011

Source: Statistics Canada, *Survey on Financing and Growth of Small and Medium Enterprises*, 2011.

One of the reasons why SMEs cannot effectively obtain loans is that between 2007 and 2011, 90-day delinquency rates for small businesses reached only 0.96%, compared with only 0.22% for medium-sized companies (Industry Canada 2013), and it implies that lending money to SMEs is riskier. This can partly explain why SMEs have lower approval rates and lower authorized-to-requested ratios.

In general, SMEs have different levels of financing difficulties depending on reasons such as the number of years of establishment, the number of employees. As a pillar of the Canadian economy, SMEs have a vital role in future economic growth. Solving the problem of financing difficulties is the key to the development of SMEs.

1.3 Difficulties to access traditional financing

SMEs contribute a lot to the Canadian economy, but face financing difficulties. Lots of researchers have mentioned that banks can be the main lenders for SMEs, and many reasons make it difficult for SMEs to borrow from banks. After communicating with some bank employees, we can conclude that the credit rating of an enterprise is one of the important criteria for determining whether a borrower can obtain loans from a bank or not. Low credit rating scores of enterprises in banks will restrict SMEs to have access to loans from banks. Normally, low credit rating scores are caused by following reasons:

1.3.1 Information asymmetry

Stiglitz and Weiss (1981) first use the theory of information asymmetry to explain the phenomenon of credit rationing: there is information asymmetry between banks and borrowers, and the requirements of raising banks' interest rate and collateral will lead to the issue of moral hazard and adverse selection from borrowers which offset the expected level of return of the bank, so the bank will set the interest rate and collateral requirements at a moderate level. At this time, the demand for funds exceeds the supply. Some of the indifferent loan applicants can obtain loans, and others are not available to loans, even if they are willing to pay higher interest rates, or no matter how abundant the loan supply, there will always be some people who are not eligible for loans under any conditions. This is credit rationing.

Due to the imperfect financial system, small enterprises are unable to provide audited financial statements, and they lack qualified collateral and guarantees. The information asymmetry between SMEs and banks is more severe than that of large enterprises. Therefore, the availability of small business loans is relatively low (Nakamura 1994).

1.3.2 Lending technology theory

Berger and Udell (2006) noted that in many cases, *a secondary information source, screening/underwriting procedure, contract feature, or monitoring mechanism is used, but they distinguish the technologies based on the primary foundations of the lending decisions. Thus, a credit score may be used as secondary information or collateral may be used as a secondary source of repayment, but the lending technology would still be relationship lending if the lending decisions are primarily based on soft information gathered over the course of a relationship.*

Berger and Frame (2007) also noted that lending technology consists of transactional lending technology and relational lending technology. Transactional lending technology relies on "hard" information, while relational lending technology depends on "soft" information. The "hard" information mainly includes the financial statements of the business, the value of the mortgaged asset, the credit rating of the business or the bank credit history. Without contacting the borrower, it is easy to be able to collect, quantify, easily verify, pass then to make loan decisions. Compared with "hard" information, "soft" information requires lenders to have contact with business owners, managers or third parties associated with the business, and it takes much time to optimize, verify and communicate. The "soft" information mainly collects information, including the ethical quality of business owners, the satisfaction of employees and customers, and the economic exchanges between enterprises and upstream and downstream partners. Based on its characteristics, the evaluation of "soft" information from financial institutions may vary. When the commercial banks make loan decisions, both "hard" and "soft" information should be considered.

SMEs are challenging to provide both. For example, Chinese SMEs in labour-intensive industries lack physical assets that can be used for mortgages, and it is hard for commercial banks to measure their "soft" information. It causes the commercial banks to be reluctant to make loans to them (Shi and Yang 2009).

1.3.3 Financial structure theory

Numerous studies have shown that large financial institutions have a comparative advantage in collecting and processing “hard” information (Stein 2002), and they mainly tend to provide loans and financial services to large enterprises; while based on geographical advantages that small financial institutions own, small financial institutions have more advantage in collecting "soft" information from SMEs and usually use relational lending technology to provide loans to small businesses (Berger and Udell 2002).

Berger and Udell (1995) argue that the size of the bank, the complexity of the organization is inversely related to the number of financial services provided to small businesses, because small business credits need to collect a large amount of information, mainly based on relational financing, but according to Williamson (1967), to provide certain product or expand the geographical area, large banks need to increase the banking organization and increase the level of management (including horizontal management and vertical management), and the scale of management will be uneconomical. Highly monopolized, centralized banking structure tends to have greater credit support for large enterprises, and credit rationing for small businesses and increased competition will improve the financing of small businesses.

1.3.4 Financial infrastructure theory

Financial infrastructure refers to the rules and conditions that affect the ability of financial institutions to provide SMEs with credit, and financial infrastructure includes the information environment, legal environment, regulatory environment and taxation system, all of which affect the legitimacy and profitability of different credit technologies that financial institutions can adopt (Berger and Udell 2006). The regulatory environment indirectly restricts the availability of SME credit by influencing the underlying financial market structure.

The information environment usually examines the reliability of accounting information and the degree of information sharing. The former mainly depends on strict accounting standards and the existence of a trustworthy independent accounting firm, while the latter can effectively reduce credit time, cost and credit risk (Miller 2003), and the presence of third-party information transactions can increase the availability of credit (Kallberg and Udell 2003).

The regulatory environment includes various regulatory regimes for financial institutions, including loan restrictions, thresholds for entry, and direct control of state-owned financial institutions. The tax system will affect the credit technology adopted by financial institutions. Berger (2006) believes that the implementation of the risk-based capital requirements of the Basel II proposal means that implicit tax imposed on SME credits makes banks carry out small business loans and more use of transactional credit technology. Comprehensive promotion of the formation of an SME regulatory environment can promote a complete and stable SME financing support system and operating mechanism.

1.4 Characteristics of FinTech Financing

The financial industry is not always reputed with user experience, transparency, and innovation. FinTech creates apps for start-up companies and is normally user-friendly for user experience, simplicity, and convenience. FinTech's are companies that combine technology and financial aspects for small business models. Within the last few years, FinTech companies have attracted attention due to their reputation for challenging incumbent financial institutions, including the traditional banking model (Boskov 2018). SMEs traditionally seek assistance to local banks for loans or seek traditional investors. If companies requested credit cards, they require accounts with larger credit providers. Strict banking regulations impacted the growth of small and medium enterprises due to limited access to finance and a tight lending environment (Temelkov and Gogova Samonikov 2018). But FinTech companies are providing alternative financing models by revolutionizing how small start-up businesses may enter their respective markets. FinTech enables crowdfunding, mobile payments, and money transfer services (Matthew 2017). Financial technology, or FinTech, is now rendering easier financial transactions and financing since the 2008 economic crisis (Arner et al. 2016). Technological innovation and developments have changed the way of managing finance by providing financial stability and market integrity to the customers.

FinTech apparently has brought the banking experiences the changes in the last decade overriding the traditional banking systems in the previous 200 years. One of the most prominent aspects of FinTech is natural mobility, which is one of the pioneers for banking simplicity. Uniformed users can perform their financial operations much differently than traditional banking means. FinTech apps provide users access to relevant information regarding their transaction history. Its popularity creates urgent needs for domestically developed tools providing easier access than traditional banking needs did. Mobile applications need to be perfectly integrated with many platforms where FinTech focuses on many sectors implying applications can be designed for customers seeking variable solutions. Mobile payments and financial resource management are among the most demanded requirements and features (Yoshida 2019). The applications are to be compatible with each other to communicate and recognize exchanged information. The numerous applications and functions are integrated into single entities mandatory for individual and business customers. The application plays an important role across different platforms. Hence the business integration factors are expected when customers need and expect them. The basic needs of customers have entirely changed with the introduction of Fintech, so applications still need to respond to customers adequately (Magnuson 2018). For existing functionalities, improvements are necessary, especially if the existing is already housing a successful application that customers expect to continue with positive experiences. In this manner controlling and operating expenses, savings, and obtaining loans becomes an easier task for customers. Another important aspect of FinTech services is the user interface while using the application. Adding the latest functions that develop and expand from the existing functions is a necessary ongoing tool for users to continue learning.

FinTech's use of artificial intelligence verifies its innovation. It has a built-in processing algorithm that interacts with the user. FinTech also has monitoring tools that detect inappropriate customer behavior and the means to counteract this type of behavior (Fenwick and Vermeulen 2017). Besides bringing cost savings, reliance on FinTech services allows for the creation of new business models not inherently prominent in the current financial ecosystem due to technical debt and obscurantism or cost prohibition. Micropayments between users, insurance or loan services offered directly between small business users that bypass rating agencies are only the beginning of fully launching FinTech (Saksonova and Kuzmina-Merlino 2017). Innovation arrives in the form of digital currencies and is also entering certain government levels. In fact, some countries are attempting to use secured digital currencies to be implemented on national levels as part of a FinTech initiation. Its secure information storage capabilities enable information to be time-stamped, so data cannot be changed or manipulated. These types of security products are known as blockchains, which is a recent innovation not yet fully adopted by small businesses (Wright 2016). By providing banking information with such security, these products can be exchanged to cover costless automated algorithms and protocols as opposed to costly security protocols under conventional regulations.

FinTech's personalization feature provides an understanding of customer wants and needs, which has been a basic principle governing global marketing. This conceptual and practical model is not too diverse from mobile banking technology and FinTech services. For FinTech to be a credible service to be relied upon, its apps need to be personalized for small businesses to be in control of their financing. Typical banking, on the other hand, often ignore small business needs to be discernible. This is the reason why FinTech start-ups can serve as the answer to small business wants and needs (Omarova 2019). SMEs can find themselves adaptable to the range of products when searching for new business ventures. FinTech apps enable correspondence between the sold products and the small business clients that require them. The individualization process finds FinTech's app integration with personal applications, including social media apps, and establishes magnified relationships with small business users (Lee 2017). Finally, financial operations are more adaptable and bring FinTech companies and clients together with positive relationships.

FinTech services have revolutionized initiatives that are conquering the financial world. Application developers will find FinTech initiatives opportunistic regardless of the software created to provide financial solutions or invest in integrations options for external applications. Even if FinTech services can be more costly; the current period of investing will bring a higher return on investment (Wonglimpiyarat 2017). If for some reason, FinTech apps suffer any drawbacks or downturns, SMEs need to understand and follow the regulations to reuse the existing apps to their advantage.

Chapter 2

Literature review

2.1 Web mining

Web mining is an application of data mining, the technique of discovering and extracting useful information from large data sets or databases (Mughal 2018). Web mining, therefore, can be defined as to discover or extract useful information from the web (Cooley et al. 1997).

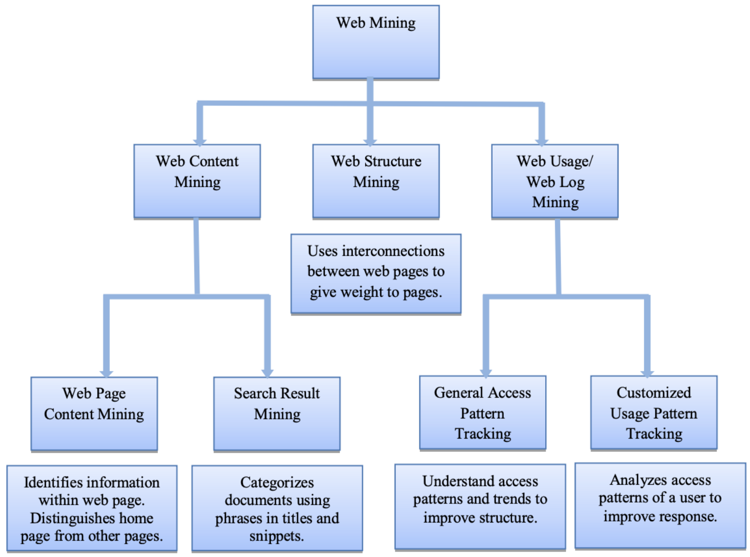
According to the different types of data (Mughal 2018) used in web mining, it can be divided into the following three categories, and they are web content mining, web structure mining and web usage mining.

2.1.1 Web structure mining

Web structure mining focuses on the hyperlink structure of the web (Costa and Zhiguo 2005a).

I-Hsien Ting (2008) defines web structure mining as a technique to analyze and explain the links and structure of websites, for which graph theory is usually the main concept and theory used. Besides, the extraction of the structure of websites is always essential in this research area (Fu et al. 2007). To avoid wrong conclusions, which was caused by applying the traditional processes and speculating that the events are independent, an

Figure 7: Web Mining Taxonomy



appropriate solutions can handle the links which could lead to potential correlations, and then improve the predictive accuracy of the learned models (Getoor 2003). Two algorithms that have been proposed to lead with those potential correlations: HITS (Kleinberg 1998) and PageRank (Page et al. 1999).

2.1.2 Web usage mining

Web usage mining is the application of data mining techniques to discover usage patterns from Web data to understand and better serve the needs of web-based applications (Srivastava et al. 2000).

M.G. da Costa and Zhiguo Gong (2005b) also thinks that web usage mining focuses on techniques that could predict the behaviour of users while they are interacting with the World Wide Web. Web usage mining collects the data from Weblog records to discover user access patterns of web pages.

2.1.3 Web content mining

Inamdar et al.(2008) and Kosala et al.(2000a) think that web content mining analyzes content on the web such as text, images videos etc. Web content mining as an application of data mining techniques to content published on the internet is used as HTML (semi-structured), plaintext (unstructured), or XML (structured) documents (Srivastava et al. 2002, Kosala and Blockeel 2000b). Web content mining can be further divided into four categories based on the scope of the usage.

Anurag mumar and Ravi Kumar Singh (2016) divided the scope of the usage for web content mining into four different categories, and they are multimedia mining, unstructured text mining, structured mining and semi-structured text mining.

- In their paper (Kumar and Singh 2016), Multimedia data mining can be defined as the process of finding interesting patterns from media data such as audio, video, image and text that are not ordinarily accessible by basic queries and associated results.
- The unstructured text mining, structured mining and semi-structured text mining are related to natural language processing (NLP), and it can be seen that recent natural language processing (NLP) is one of the main technology that used in this area.

Sukhija (2015) thinks that natural language processing helps to understand how to manage unstructured data over the machine processing platform using various NLP techniques along with web content mining.

2.1.4 Raw data extraction

The advancement in information technology and development in databases has caused great difficulty in extracting data from the huge databases. These enormous databases result in big data, which requires efficient-explorational abilities and appropriate handling to make the best use of the available data (Ibrahim et al. 2018). There are different methods used for data extraction and exploration, including APIs, commercial sources, web crawlers, and public sources.

2.1.5 Method 1: Web Crawlers and Public Sources

With the development of internet technology, a huge list of web pages is getting added, which changes the available information every day. To simplify this process and provide ease to the users, search engines are used for the required information extraction. Search engines consist of a principal part known as web crawlers, which are used for data mining and extraction. Web crawlers are basically computer programs, which are used for browsing the web worldwide by operating an orderly, automated or methodical manner. They are further used to help users for web navigation by using search engines. Moreover, they are significant in data collection, management, handling, and exploration with the growing internet. Also, they enable the users to look for the sources with the help of various hypertext links (Amudha and Phil 2017).

A study conducted by Thelwall (2001) highlights the design and significance of a web crawler used in data mining. According to him, web content is widely researched globally. For the web page processing, various computer programs are used, and web crawler is one of them. It is used for data mining and fetching web pages for better analysis. Therefore, this study is conducted to propose a distributed system, which is used for efficiently using computing resources and helping information managers to use effectively extracted data without using any expensive types of equipment (Thelwall 2001). Another study conducted by Singh & Varnica (2014) emphasizes the significance of a web crawler fused for data extraction. According to the study, a massive increase in the use of the internet has been observed worldwide. Users use different hyperlink texts to search on the web engine. Due to which, web crawlers are introduced. These are used for navigating purposes on the web and generally known as text search engines. In addition, research activities can also be conducted by using web crawlers. There are different types of web crawlers that are usually dependent on the types of search engines being used, including the Breadth-First Crawler, Incremental Web Crawler, Form Focused Crawler, Focused Crawler, Hidden Web Crawler, Parallel Crawler, and Distributed Web Crawler (Ahuja et al. 2014).

2.1.6 Method 2: Commercial Sources

Another method used for data mining and extraction includes commercial sources, which are given as web scrapers. These web scrappers are expensive commercial software. Also, some of the web scrapers are available free of charge, having limited accessible features to work with. In addition, web scraping is known as the extraction of web data or web harvesting. It is used for extracting web information available on different websites. For this purpose, users use different web scrapers for worldwide web exploration by using embedded web browsers or hypertext transfer protocol. It is also connected to web indexing. It is a technique used to extract information and used by various search engines for information indexing on the web. Moreover, some web scrapers transfer the unstructured data into the structured data, which can be handled and evaluated in a spreadsheet or local database. In addition, web scrapers are used for comparing online prices, integrating web data, web mashup, web research, detecting website change, monitoring weather data, and so on. There are various software tools (commercial) available, which are used for website personalization by using web scrapers and given as Rapid miner, Orange, Weka, Knime, Oracle data mining, Rattle, and so on (Vargiu and Urru 2013).

In addition, a study conducted by Haddaway (2015) highlights the various uses of web scraping software, including grey literature search. According to the International Conferences on Grey Literature (Grey Literature International Steering Committee 2006), grey literature is defined as follows: "*Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing, i.e. where publishing is not the primary activity of the producing body*." According to the study, grey literature research requires several sources for extensive and efficient searching. For this purpose, web scraping is used for patterned data extraction available on the website. These commercial web scrapers have also been created by private businesses. In addition, they provide various benefits for grey literature research. By using web scrapers, the resource efficiency and transparency of grey literature research can significantly be improved. Also, the study (Haddaway 2015) highlights various available software used for web scraping.

2.1.7 Method 3: Application Programming Interface (API)

API is a user interface that is used for performing different functions by over-topping the application. Globally, users use various scripts and other methods for extracting and scraping data, but with the help of API, users can extract data on their own systems by using a user-friendly and understandable interface with no complex functions and codes. In addition, it enables you to extract the required data by integrating the API with the applications used(Abdulrahman et al. 2013). Detailed documentation is available for an easy and user-friendly understanding of the use of the API. With the help of API, users can extract structured and clean data with no human effort (Shen et al. 2019). There are various applications of APIs, including display content in another application, embedded content, and extract data. This user-friendly interface enables the users to extract every kind of data, including scanned pdfs, notes, social media applications, images, and so on. In addition, these tools are further used for research purposes of extracting the data more accurately and programmatically (Cuesta et al. 2014).

A study conducted by Lomborg & Bechmann (2014) enlightens the use of APIs for collecting data on social media platforms. According to this study, APIs are developed by social media companies for providing data by using APIs to have detailed research on social media. It is basically an add-on to an application and works as a back-end interface. It is further used for data collection to have a thorough empirical analysis of the data. In addition, the study provides the benefits and challenges offered by APIs for a detailed analysis of social media research, including both qualitative and quantitative.

2.2 Text mining

2.2.1 The definition of text mining

Text mining has become an important area of research. Pons-Porrata et al. (2007) gave a definition of text mining, and text mining refers to the process of extracting interest information and non-retrieving information or knowledge from unstructured text collections.
Text data mining is a specific research field in data mining, and its main task is to discover potential rules and trends in massive texts. Unlike the narrow sense of data mining, text mining has a process of transforming natural language text into structured data that can be processed by computers. The main purpose of this purpose is to extract metadata representing text features (Xu 2017).

2.2.2 Text mining techniques

In Shilpa Dang's (2014) paper, text mining is a multidisciplinary field involving information retrieval, text analysis, information extraction, categorization, clustering, visualization, data mining and machine learning. The basic steps of text mining can be divided in five steps.
The first is to extract unstructured data from the chunk of information. The second converts the information received into structured data. The third is to identify patterns from the structured data. The fourth step is to analyze the patterns. The fifth extract valuable information and stores it in a database.
From another perspective, we can summarize these steps as follows: Information Extraction (IE), Summarization, Clustering, Categorization, Topic Modelling and their applicability.

2.2.3 Corpora

Corpora are collections of related documents that contain natural language and are used to cope with updated text data from information retrieval. The commonly used corpus includes Brown corpus, Wikipedia corpus or Cornell movie dialogue corpus, and so on (Bengfort et al. 2018a). Although there are many open-source corpora. In some specific fields, there is a tendency to use the specific corpus to process data to avoid the ambiguous situation.

2.2.4 Data preprocessing

The bag-of-words method (BOW):

The bag-of-words method can provide an overview of the dataset. The bag-of-words model (BOW) can convert a sentence into a vector representation. It does not consider the order of words in the sentence, only the frequency of occurrence of the words in the sentence (Kanakaraj and Guddeti 2015). For example, in a huge document set D, there are a total of M documents, and all the words in the document are extracted to form a dictionary containing N-words. Each document can be converted to an N-dimensional vector via the BOW model. So the N*M matrix is created. The BOW model has two disadvantages (Kanakaraj and Guddeti 2015). First, it does not consider the order of words, and second, it cannot reflect the keywords of a sentence.

Stemming and lemmatization:

Stemming is defined to reduce words to a character response likely related to their linguistic root (Singh and Gupta 2017). Stemming uses a series of rules (or a model) to slice a string to a smaller substring. The goal is to remove word affixes (particularly suffixes) that modify the meaning.

Lemmatization uses a dictionary to look up every token and returns the canonical "head" word in the dictionary, called a lemma (Bengfort et al. 2018b). Because it is looking up tokens from ground truth, it can handle irregular cases as well as handle tokens with different parts of speech.

Stopwords:

To reduce the dictionary size, it is necessary to compile a list of stopwords and remove those stopwords from the dictionary. These words almost never have any predictive capability, such as articles *a* and *the* and pronouns such as *it* and *they* (Akaichi et al. 2013).

TF-IDF:

TF-IDF is used to calculate the relative frequency or rareness of tokens in the document against their frequency in other documents. The formula used is tf (t, d) as a boolean frequency or the count. N represents the number of documents, and nt represents the number of occurrences of the term t in all documents (Kobayashi and Aono 2008). The formulas are as follows:

Figure 8: TF-IDF formula

$$tf(t, d) = 1 + \log f_{t,d}$$

$$idf(t, D) = \log 1 + \frac{N}{n_t}$$

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$

Source: Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning

Single Value Decomposition (SVD):

Typically the word by context matrix is very large and typically sparse. Single Value Decomposition (SVD) is used to reduce the number of dimensions without great loss of descriptiveness, and it is the underlying operation in a number of applications including statistical principal component analysis, text retrieval, pattern recognition and dimensionality reduction, and natural language understanding (Maletic and Valluri 1999).

2.2.5 Clustering for text similarity

Clustering algorithms aim to discover latent structure or themes in unlabeled data using features to organize instances into meaningfully dissimilar groups (Bengfort et al. 2018b). The behaviour of unsupervised learning methods is fundamentally different; instead of learning a predefined pattern, the model attempts to find relevant patterns a priori.

2.2.6 Distance metrics

We also need to consider "distance metrics". A corpus is transformed into feature vectors, and a clustering algorithm is employed to create groups or topic clusters, using a distance metric such that documents that are closer together in feature space are more similar. New incoming documents can then be vectorized and assigned to the nearest cluster.

Bengfort et al. (2018b) consider documents as points in space, where the relative closeness of any two documents is a measure of their similarity.

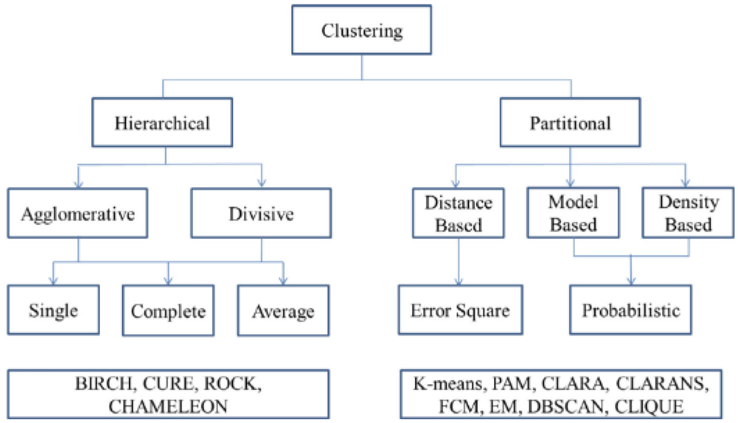
Some distance methods can be used, such as Euclidean distance. Euclidean distance is a standard metric for geometrical problems. It can be defined as the ordinary distance between two points and commonly used in two- or three-dimensional space. Euclidean distance is commonly used with the k-means algorithm.

In the case of texts, cosine distance is effective. It is defined from the cosine of the angle between vectors. *W*hen documents are represented as term vectors, the similarity of two documents corresponds to the correlation between the vectors (Huang 2008). When dealing with distances, the cosine similarity will not consider document length and is not biased by the asymmetry of information.

2.2.7 Clustering methods

A bunch of algorithms can be used when dealing with clustering problems. From figure 9, we can see that partitive clustering and hierarchical clustering are two main approaches.

Figure 9: Taxonomy of clustering approaches



Source: A review of clustering techniques and developments (Saxena et al. 2017a)

Partitive clustering:

After quantifying the similarity of two documents, unsupervised methods are used to find similar groups of documents. In partitive clustering, partitive methods partition instances into groups that are represented by a central vector (the centroid) or described by a density of documents per cluster. Centroid represents an aggregated value (e.g., mean or median) of all member documents and is a convenient way to describe documents in that cluster. Among these methods, *K-means algorithm is one of the best-known,bench marked and simplest clustering algorithms, which is mostly applied to solve the clustering problems*(Saxena et al. 2017b). K-means clustering algorithm starts with an arbitrarily chosen number of clusters, k, and partitions the vectorized instances into clusters according to their proximity to the centroids, which are computed to minimize the within-cluster sum of squares. The optimization of k means includes an experiment with different values of k (Bengfort et al. 2018b).

Hierarchical clustering:

The hierarchical clustering algorithm is an algorithm that attempts to establish a hierarchy of clusters. It has two strategies, the first is agglomerative, and the other is divisive.

Agglomerative is a bottom-up approach. It initially assumed that each observation was an independent cluster, and then merged two adjacent layers on each layer. Each cluster eventually reaches all the highest levels, and all observations belong to the same cluster. In contrast, divisive is a "top-down" approach, which initially assumes that all observations belong to the same cluster, and then separates the observations that are relatively similar in the middle of each cluster to the next layer, and all observations belong to different clusters (Sasirekha and Baby 2013).

2.2.8 Text mining application and framework

Text mining has been widely used in many fields. The most common areas of text mining applications include (Bolasco et al. 2006):

- Publishing and media
- Telecommunications, energy and other service industries
- Information Technology Department and the Internet
- Banking, insurance and financial markets
- Political institutions, political analysts, public administration and legal documents
- Pharmaceutical and research companies and health care

It can be seen that text mining has a wide range of applications in many fields. In business environment applications, text mining plays an important role in helping organizations and businesses analyze customers and competitors. It provides insight into the business and provides information, improve customer satisfaction and gain a competitive advantage (El Zahra et al. 2010).

Chapter 3

Methodology

3.1 Web Mining

Web scraping is a powerful tool for working with data on the web. With a web scraper, we can get a large corpus of text or quantitative data to work with.

After many experiments, we can design our web scraper based on following steps. First, we need to understand the structure of a web page. The second is to filter the useful parts as needed and download the corresponding web contents. The third is to parse the downloaded webpage. The forth is to store useful information.

Before the data collection, we should know the data collection environment and tools involved:

Figure 10: The Environment and Tools on Data Collection

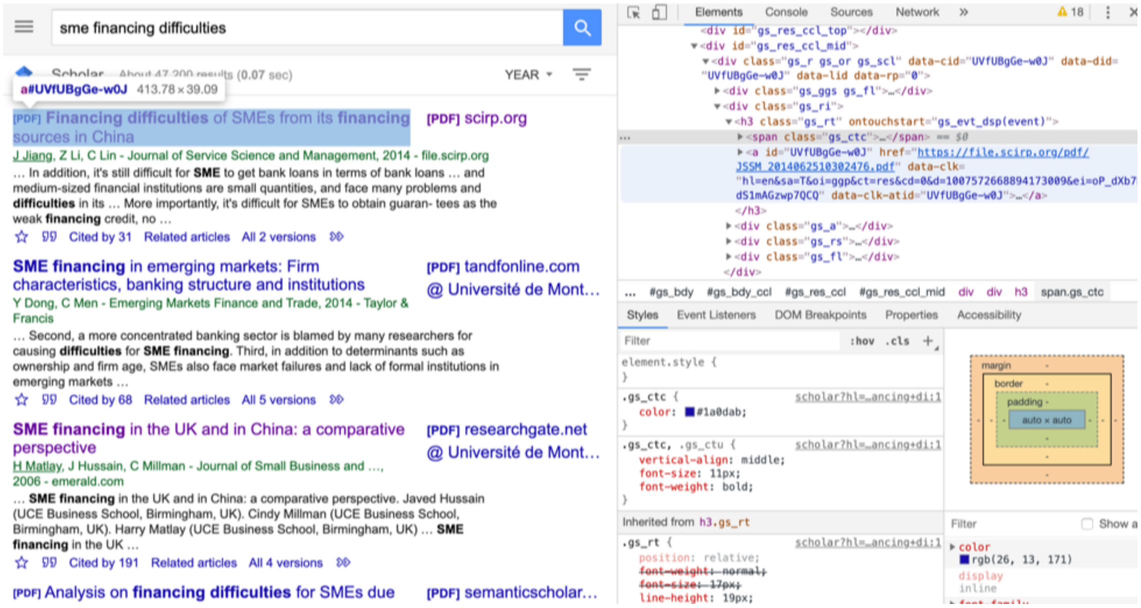
The Environment and Tools on Data Collection

Web-based user interface	JupyterLab Python 3.7
Web Scraping tools	requests
Processor	1.7 GHz Quad-Core Intel Core i7
Memory	8 GB 2133 MHz LPDDR3

3.1.1 The structure of the web page

When we click on 'inspect' on any web page, and we can see the structure of HTML.

Figure 11: HTML inspection



In fact, in addition to HTML, two other components build colourful/multifunctional web pages, and they are CSS and JavaScript. However, our goal here is to extract the titles of articles and the URLs of them. The understanding of CSS and JaveScript is out of this scope, so we will not discuss about these two elements now.

In the process of designing a crawler, the most troublesome step is to analyze the element nodes in the web page and collect relevant information. This process requires a clear interpretation of the HTML file structure, which can accurately locate the required dom nodes and the required data content in the complex dom structure.

By observing the structure of HTML, we can see that HTML consists of a series of elements. HTML elements tell the browser how to display the content. HTML elements are represented by tags. HTML tags label pieces of content such as "heading", "paragraph", "table", and so on. Browsers do not display the HTML tags but use them to render the content of the page. By querying the source code of the page, we can see the elements structure information of the entire page. From figure 11 and 12, relevant data such as the title of the useful articles and the text links of the secondary page need to be obtained.

Figure 12: The needed elements from web inspect

The needed elements from web inspect

Name	Tag and attributes
Title	Under "h3", attrs={"class": "gs_rt"} link['href']
URL	Under "h3", attrs={"class": "gs_rt"} link.text

3.1.2 The design of web crawlers

We first need to analyze the URLs of the website. For the processing of dynamic request data, we can repeatedly try to get a simple answer after grabbing the request URL. For example, in the google scholar's page crawling process, we need to know the specific request parameters of the page data URL on the next page. By manually turning the page multiple times and comparing the changes in the URL parameter, the page-turning parameters are represented.

After comparison, we found that the webpage except the first page is a little different from the other pages, but the other URLs are regular: only the number after the last "start =" is changing, and it is incremented by 10. We first assume that the webpage of the first page is "start = 0". From figure 13 to figure 15, we presented the URLs with changing patterns.

Figure 13: First page URL

scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Fintech+SME+solutions&oq=Finte

Figure 14: Second page URL

scholar.google.com/scholar?start=10&q=Fintech+SME+solutions&hl=en&as_sdt=0,5

Figure 15: Third page URL

scholar.google.com/scholar?start=20&q=Fintech+SME+solutions&hl=en&as_sdt=0,5

Based on the URL patterns we found, we can create codings to extract the webpages we want.

Many methods can be used to extract website information and for simplicity, we used "requests" function to extract website information.

We can summarize the process when plugging into our keywords. The process of information extraction is as follows:

- The first step is to select the initial interface for crawling and fill in the value of the initial page in the URLs' property, which is an empty set.
- The second step is to determine search names we need for google scholar. For the thesis, we will use "SME financing difficulties" and "Fintech SME solutions" as our keywords. We need to filter 50 pages for titles and URLs. For every webpage, there are ten different titles and URLs we can use. Our goal is to extract 1000 URLs and titles to build our dataset. Based on the patterns we found for the links, we need to build a loop to fetch important titles and URLs information.
- We will use the 'beautifulsoup' function to parse the content obtained from 'requests" function.
- We will also set a dwell time in 1s for the web page.
- The crawled contents will be stored in the form of CSV as shown in figure 16.

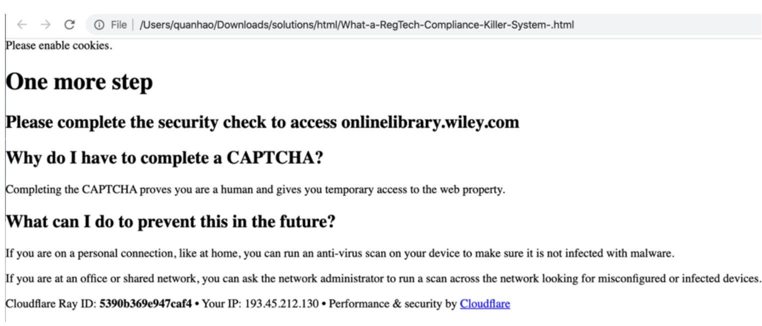
As shown in figure 16, it can be seen that the URLs are PDF files and HTML files. When we randomly clicked in HTML files, and we found that there is an encrypted problem, as shown in figure 17.

To address the problem shown in figure 17, we will have separate HTML files and PDF files, and we will only keep PDF files to continue our research.

Figure 16: Titles and URLs extracted from the websites

title	url
Financing difficulties of SMEs from its financing sources in China	https://file.scirp.org/pdf/JSSM_2014062510302476.pdf
SME financing in emerging markets: Firm characteristics, banking structure and institutions	https://www.tandfonline.com/doi/abs/10.2753/REE1540-496X500107
SME financing in the UK and in China: a comparative perspective	https://www.emerald.com/insight/content/doi/10.1108/14626000610705769/full/html?queryID=51%2F5408313
Analysis on financing difficulties for SMEs due to asymmetric information	https://pdfs.semanticscholar.org/600e/c2e0e6df0fdaf3ecebb958259a08327886f6.pdf
SME firm performance-financial innovation and challenges	https://www.sciencedirect.com/science/article/pii/S1877042815038409
Small business financing initiatives in Ghana	https://www.researchgate.net/profile/Nicholas_Biekpe/publication/265484299_Small_Business_Financing_Initiatives_in_Ghana_1/links/54e6f2990cf277664ff760a2/Small-Business-Financing-Initiatives-in-Ghana-1.pdf
SME development in China: A policy perspective on SME industrial clustering	http://www.eria.org/SME%20Development%20in%20China_A%20Policy%20Perspective%20on%20SME%20Industrial%20Clustering.pdf
Bank market structure, competition, and SME financing relationships in European regions	https://link.springer.com/article/10.1007/s10693-009-0060-0
The factors of affecting financing and countermeasures based on asymmetric information and marginal information cost	https://link.springer.com/chapter/10.1007/978-3-642-29084-8_104
Countermeasures for Rural Financing Difficulties in Western China under Double Dual Structure——Research Based on Evolutionary Game Theory [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-DJKX201101003.htm
Banking system soundness is the key to more SME financing	https://www.econstor.eu/handle/10419/106312
A Study of the Connecting Channel between SME Financing and Private Funds: Analysis and Reflection on the Fund Management System of Informal Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JXCZ201101007.htm
Economic Analysis of the Financing Difficulties for Private SME [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CQYD200906009.htm
Group Financing Model and Path Direction——Based on Game Analysis of the SME Decision-making and Rational Breakout [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CJKX200905007.htm
Financing System Study of SME under New Institutional Economics [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-XTGS201106020.htm
The financing of technology-based small firms: A review of the literature	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=759544
Economic Analysis on Financing Credit Chain of Medium and Small Enterprises [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-HNJR200706010.htm
The Routing Selection of the SME's Financing Platform Construction [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-HLJC201004005.htm
Analysis on financing difficulties of small and medium-sized enterprises in China and corresponding countermeasures	https://pdfs.semanticscholar.org/42ac/7e7cc84ca991a7d605a8e569b03d5a140b06.pdf
Blockchain, bank credit and SME financing	https://link.springer.com/article/10.1007/s11135-018-0806-6
Loan rate pricing of SME financing based on agent-based computational finance approach	https://www.sciencedirect.com/science/article/pii/S1874865110600877
WORKING CAPITAL FINANCING PREFERENCES: THE CASE OF MAURITIAN MANUFACTURING SMALL AND MEDIUM-SIZED ENTERPRISES (SMEs).	https://www.researchgate.net/profile/Kesseven_Padachi/publication/268685548_Working_capital_financing_preferences_The_case_of_Mauritian_manufacturing_small_and_medium-sized_enterprises_SMEs/links/547349b50
Different Guarantee Institutions and Their Effects on the Financing of Small and Medium-Sized Enterprises (SMEs)[J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JJYJ200509006.htm
Interest Rate Marketization, Relationship Lending and SME Financing	http://en.cnki.com.cn/Article_en/CJFDTotal-XCXB201404006.htm
Impediments to SME access to finance and credit guarantee schemes in China	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=486204
Fisheries SMEs Financing Difficulties and Countermeasures	http://en.cnki.com.cn/Article_en/CJFDTotal-NYJG201206019.htm
The Analysis of civil Financing theory: the scope, the Comparison and the InstitutionalTransition [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JRYJ200408012.htm
Analysis on SME Financing Difficulty, Private Lending Difficulty and Institutional Reform [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-DJGL201302016.htm
Financing the growth of SMEs in Africa: What are the contraints to SME financing within ECOWAS?	https://www.sciencedirect.com/science/article/pii/S1879933717300362
Comparison analysis of small and medium enterprise financing structure—— questionnaire basing on Jiangshu province's small and medium enterprises [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TAJJ200803015.htm
Financing canadian SME exporters	https://www.ic.gc.ca/eic/site/061.nsf/vwapj/Financing_Canadian_SME_Exporters_Eng.pdf/$FILE/Financing_Canadian_SME_Exporters_Eng.pdf
Four proposals for improved financing of SME development in ASEAN	http://www.academia.edu/download/33295754/finance_sme.pdf
Research on SME Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-SGXB200905011.htm
Financing small businesses in France	https://www.econstor.eu/handle/10419/44830
Causes for SME's Financing Problem in China and its Countermeasures [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-YSXS200403020.htm
Financing high-tech SMEs in China: A three-stage model of business development	https://www.tandfonline.com/doi/abs/10.1080/08985620903233937
Developing the capital market to widen and diversify SME financing: the Korean experience	https://www.asean.org/storage/images/archive/22633-8.pdf
A Study on the Structure of Financing Demand of Small and Medium Enterprises and the Counter-measures [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CSJR200908011.htm
An Analysis of perfection of SME Legal System in China [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TYDX200703015.htm
Changes in Interest Rate, Shadow Banking and SME Financing	http://en.cnki.com.cn/Article_en/CJFDTotal-CSJR201307007.htm
An empirical evidence of small business financing in China	https://www.emeraldinsight.com/doi/abs/10.1108/01409170810920666
On The Financing Predicament and Strategy of SME [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JGGB200902012.htm
The Research on Inventory Pledge in SME Financing	http://search.proquest.com/openview/3efdb691503fd69572deb9f49b8fdb82/1?pq-origsite=gscholar&cbl=2028701
Relational Financing in Small and Medium-sized Enterprises: empirical study and interpretation [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-SXCJ200402021.htm
Bank competition, government intervention and SME debt financing	https://www.emeraldinsight.com/doi/abs/10.1108/CFRI-02-2017-0007
Japanese banking regulations and SME finance under the global financial crisis	http://jsmeweb.org/jjmfe/pdf/jjmfe_2013august_4.pdf
Information, Informal Finance and SME Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JJYJ200507003.htm
Liaonmg small and middle enterprise financing difficulty forming reason and countermeasure research [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TAJJ201209065.htm
An Analysis of Financing Difficulty of Small and Medium Enterprises in the New Rra [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-ZGSG201205012.htm
On Legislation of Private Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-NFZJ201201010.htm
Research on Financing Cost of Small and Medium-Sized Enterprises by Internet Finance	https://www.scirp.org/html/7-1761631_80477.htm
Relationship lending and SME financing in the continental European bank-based system	https://link.springer.com/article/10.1007/s11187-008-9129-7
Analysis on financing problems for small and medium-sized enterprises	https://ieeexplore.ieee.org/abstract/document/6013702/

Figure 17: HTML result problem


File | /Users/quanhao/Downloads/solutions/html/What-a-RegTech-Compliance-Killer-System-.html

Please enable cookies.

One more step

Please complete the security check to access onlinelibrary.wiley.com

Why do I have to complete a CAPTCHA?

Completing the CAPTCHA proves you are a human and gives you temporary access to the web property.

What can I do to prevent this in the future?

If you are on a personal connection, like at home, you can run an anti-virus scan on your device to make sure it is not infected with malware.

If you are at an office or shared network, you can ask the network administrator to run a scan across the network looking for misconfigured or infected devices.

Cloudflare Ray ID: **5390b369e947caf4** • Your IP: 193.45.212.130 • Performance & security by Cloudflare

3.1.3 The classification of URLs

By observing the URL files, it can be seen that under the title, the file form is mainly divided into two categories (illustration): PDF and HTML.

Figure 18: URLs without classification

https://core.ac.uk/download/pdf/144468918.pdf
https://www.biomedcentral.com/openurl?doi=10.1186/s40854-016-0037-6
https://www.cepal.org/sites/default/files/events/files/2-1._alejandro_chavez_0.pdf
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215243
https://ojs.amhinternational.com/index.php/jsds/article/view/2478
https://link.springer.com/chapter/10.1007/978-3-319-54666-7_4
https://link.springer.com/chapter/10.1007/978-3-030-02831-2_5
https://www.sciencedirect.com/science/article/pii/S0007681317301246
https://idl-bnc-idrc.dspacedirect.org/handle/10625/57158
https://link.springer.com/chapter/10.1007/978-3-319-66119-3_6
https://pdfs.semanticscholar.org/0c16/cfe3da4583db360eb546d3a63fa597aa6856.pdf
https://www.emerald.com/insight/content/doi/10.1108/978-1-78973-545-120191018/full/EthisKapital.com

The goal of establishing classification is to reconstruct the two forms of CSV according to the two forms of PDF and HTML.

From figure 18, it can be seen in the URL that all secondary pages that end with PDF are all PDF, and all other pages that do not end with PDF are all directed to the next web. One idea of classification is to distinguish URLs with regular expressions, which can find pages of PDFs through regular expressions. Pages that are not PDFs are filtered into another CSV through an 'if loop'

By observation, some redundant separartors and symbols need to be normalized. Created in the for loop to regulate the path.

After deciding the type of URL, downloading the file is divided into 'SME financing difficulties' type and 'Fintech SME solutions' type according to the difference of google search. There are four types of documents downloaded from the classification. From figure 19 to figure 22, we present the results that extracted from the website.

Figure 19: PDF files of "Fintech SME solutions"

pdf-scholar_files_solution

title	url
FinTech in Germany	https://link.springer.com/content/pdf/10.1007/978-3-319-54666-7.pdf
INSTITUTIONALIZATION OF SME'S IN TURKEY AND A COMPUTATIONAL FINANCIAL MANAGEMENT SOLUTION SUGGESTION	https://www.researchgate.net/profile/Hakan_Yildirim2/publication/333455021_INSTITUTIONALIZATION_OF_SME%27S_IN_TURKEY_AND_A_COMPUTATIONAL_FINANCIAL_MANAGEMENT_SOLUTION_SUGGESTION/links
Emergence of Fintech Companies in Southeast Asia—Rising Hopes of a Solution to Financial Issues—	https://www.jri.co.jp/MediaLibrary/file/english/periodical/rim/2018/68.pdf
Awareness and perception analysis of small medium enterprise and start-up towards FinTech instruments: Crowdfunding and peer-to-peer lending in Malaysia	https://www.researchgate.net/profile/Nasrul_Ghazali/publication/324038896_Awareness_and_Perception_Analysis_of_Small_Medium_Enterprise_and_Start-up_Towards_FinTech_Instruments_Crowdfunding_and_Peer-to-Pee
Future of FinTech	https://link.springer.com/content/pdf/10.1007/978-3-319-51415-4.pdf
THE POTENTIAL OF THE FINTECH INDUSTRY TO SUPPORT THE GROWTH OF SMES IN INDONESIA 35142338-8 RARAS MINERVA MANAGEMENT ...	https://core.ac.uk/download/pdf/144468918.pdf
Efforts to create a more financially-sustainable condition for SME's	https://www.cepal.org/sites/default/files/events/files/2-1._alejandro_chavez_0.pdf
Adoption of Shariah Compliant Peer-To-Business Financing Platforms by Smes: A Conceptual Strategic Framework for Fintechs in Bahrain	https://pdfs.semanticscholar.org/0c16/cfe3da4583db360eb546d3a63fa597aa6856.pdf
Digital Innovation Can Improve Financial Access for SMEs	https://t20japan.org/wp-content/uploads/2019/03/t20-japan-tf9-1-digital-innovation-improve-financial-access-smes.pdf
Sequential coherence as a determinant of open innovation: A case study on a Sri Lankan fintech solution	http://www.maco.jfn.ac.lk/ijabf/wp-content/uploads/2019/11/vol05issue1-01.pdf
FinTech Revolution	https://link.springer.com/content/pdf/10.1007/978-3-319-76014-8.pdf
The fintech challenge	http://real.mtak.hu/89231/1/Challenges_233_Kerenyi.pdf
The rise of alternative finance: how fintech companies are revolutionizing the financial service and the traditional bank financing	http://tesi.cab.unipd.it/53120/1/Dinardo_Francesca.pdf
BIG DATA AND FINTECH IN ISLAMIC FINANCE: PROSPECTS AND CHALLENGES	http://conference.kuis.edu.my/mfifc/images/e-proceeding/2018/236-244.pdf
FinTech in China's Capital Market	http://www.nomurafoundation.or.jp/en/wordpress/wp-content/uploads/2017/09/NJACM2-1AU17-03_CHINA.pdf
FINANCIAL INCLUSION: THE ROLE OF FINTECH AND DIGITAL FINANCIAL SERVICES IN INDIA	https://www.researchgate.net/profile/Dr_Vinay_Kandpal/publication/337168283_FINANCIAL_INCLUSION_THE_ROLE_OF_FINTECH_AND_DIGITAL_FINANCIAL_SERVICES_IN_INDIA/links/5dc98805299bf1a47b2f9788/FINAN
Bracing for seven critical changes as fintech matures	http://school-of-finance.ru/materials/2.5.pdf
Fintech Finance and Financial Fragility: Focusing on China	https://www.researchgate.net/profile/Robert_Van_Order/publication/321185322_Fintech_Finance_and_Financial_Fragility-Focusing_on_China/links/5b97ea7992851c4ba80fd08d/Fintech-Finance-and-Financial-Fragility-Focus
FinTech: Evolution and regulation	https://law.unimelb.edu.au/__data/assets/pdf_file/0011/1978256/D-Arner-FinTech-Evolution-Melbourne-June-2016.pdf
The Impact of Fintech On The Sustainability Of Islamic Accounting And Finance Education In Malaysia	https://www.researchgate.net/profile/Nurul_Jamil2/publication/332440091_The_Impact_of_Fintech_On_The_Sustainability_Of_Islamic_Accounting_And_Finance_Education_In_Malaysia/links/5cb56778299bf12097684d48/Th
FinTech and RegTech: Opportunities and Challenges	http://www.law.hku.hk/aiifl/wp-content/uploads/ppt/JFR-Arner_Douglas_ppt.pdf
The promise of Fintech–something new under the sun	https://www.fsb.org/wp-content/uploads/The-Promise-of-FinTech-%E2%80%93-Something-New-Under-the-Sun.pdf
FINTECH: USHERING IN A DIGITAL REVOLUTION	http://navajyotijournal.org/February_2019/FINTECH%20USHERING%20IN%20A%20DIGITAL%20REVOLUTION_Feb_2019_6.pdf
Mexico's Nascent Fintech Offers Promise, Faces New Rules	https://www.dallasfed.org/~/media/documents/research/swe/2018/swe1804c.pdf
Peer-to-peer (P2P) lending platform adoption for small medium enterprises (SMEs): a preliminary study	http://www.ijafb.com/PDF/IJAFB-2018-10-06-01.pdf
Business Agility Strategy: Peer-to-Peer Lending of Fintech Startup in the Era of Digital Finance in Indonesia	http://buscompress.com/uploads/3/4/9/8/34980536/riber_8-s4_17_h17-124_239-246.pdf
The Disruption of Fintech On Rural Bank: An Empirical Study On Rural Banks in Indonesia	https://www.researchgate.net/profile/Silvia_Anggreini/publication/336891020_The_Disruption_of_Fintech_On_Rural_Bank_An_Empirical_Study_On_Rural_Banks_in_Indonesia/links/5db93a5da6fdcc2128ebe435/The-Disruptic
Financial and banking problems facing Romania in the current and future period	https://www.researchgate.net/profile/Otilia_Manta/publication/324452083_Financial_and_banking_problems_facing_Romania_in_the_current_and_future_period/links/5d6e994192851c1b3b27654f/Financial-and-banking-prol
Competition issues in the Area of Financial Technology (FinTech)	https://www.startmag.it/wp-content/uploads/ipol_stu-1.pdf
Fintech phenomenon: a financial revolution or a new financial concern?	http://tesi.cab.unipd.it/62623/1/Saltarelli_Silvia.pdf
Fintech risk management: A research challenge for artificial intelligence in finance	https://pdfs.semanticscholar.org/cbdc/7a3429c2f67f0dcc7f14bc6e3bc811d19524.pdf
SME Financing–How To	https://nathaninc.com/wp-content/uploads/2016/05/Topic-Guide-SME-Financing-How-To.pdf
How can SMEs support growth in a low-growth region?: A longitudinal study of SME growth facilitation	https://core.ac.uk/download/pdf/50696272.pdf
Financializing authoritarian capitalism: Chinese fintech and the institutional foundations of algorithmic governance	http://financeandsociety.ed.ac.uk/ojs-images/financeandsociety/FS_EarlyView_Gruin.pdf
FINTECH FOR FINANCIAL INCLUSION	https://pdfs.semanticscholar.org/6e48/8925437965828ee1a2f6df573c675f1af2c8.pdf
SME techno-entrepreneurship: drivers and barriers in sub-Saharan Africa	https://www.elgaronline.com/downloadpdf/edcoll/9781786439062/9781786439062.00023.pdf
Chinese Fintech Companies' International Development: Motivations to Pursue International Expansion	http://www.nexgenforum.org/uploads/6/1/7/3/61732105/winter_research_trip_report_%E2%80%93_fintech_group.pdf
Gaining from digital disruption: The Thai financial landscape in the digital era	https://www.pier.or.th/wp-content/uploads/2019/09/paper5_paper_Thammarak-Vorapat.pdf
G20/OECD Effective Approaches for Implementing the G20/OECD High-Level Principles on SME Financing	https://pdfs.semanticscholar.org/cb63/39f0184dd3932f7ec37f5771e09ff9538c87.pdf
Financial Inclusion through Efficient Zakat Distribution for Poverty Alleviation in Malaysia: Using FinTech & Mobile Banking	http://conference.kuis.edu.my/icomm/5th/images/eproceeding2018/IC-002.pdf
Bank and fintech competitive dynamics and the perceived value of partnerships in an open banking market environment	https://pdfs.semanticscholar.org/6e38/ebb3cf6251f41739c5830af8c4b33f787e7b.pdf
FinTech/RegTech	https://www.iif.com/portals/0/Files/private/32370132_van_liebergen_-_machine_learning_in_compliance_risk_management.pdf
The Future of FinTech	https://srdas.github.io/Papers/fintech.pdf
Data: The Fuel of the Digital Economy and SME Growth	https://www.accenture.com/t20160901T103414__w__/us-en/_acnmedia/PDF-29/Accenture-Data-The-Fuel-Of-The-Digital-Economy-And-SME-Growth.pdf
Evaluating the Efficiency of Technology Innovation in SMEs	http://icsb.org/wp-content/uploads/2016/01/ICSBKorea-1010-abstracts-all.pdf
Eximchain: Supply Chain Finance solutions on a secured public, permissioned blockchain hybrid	https://www.eximchain.com/Whitepaper-Eximchain.pdf
FinTech: how technologies are shaping banking future	http://tesi.luiss.it/25007/1/212531_VALLI_EDOARDO.pdf
A CROSS-NATIONAL STUDY OF THE DRIVERS OF FINTECH STARTUPS: THE ROLE OF CULTURE	https://lib.ugent.be/fulltxt/RUG01/002/784/243/RUG01-002784243_2019_0001_AC.pdf
Expanding Data Collection to Streamline MSME (Micro, Small and Medium Enterprise) Lending	https://t20japan.org/wp-content/uploads/2019/05/t20-japan-tf2-15-expanding-data-collection-to-streamline-msme.pdf
Bangladesh	http://www.microsave.net/wp-content/uploads/2018/11/Bangladesh_Country_Focus_Note.pdf
Understanding Fintech and Private Equity	https://pure.uva.nl/ws/files/35456964/arnoud_w_a_boot_jeroen_e_ligterink_and_jens_k_martin_2018_understanding_fintech_and_private_equity_topics_in_corporate_finance_n_1_.pdf
How African fintech firms adapt their marketing strategies for successful intra-Africa expansion	http://gsblibrary.uct.ac.za/researchreports/2018/Hammerschlag.pdf
Platinum Sponsor	https://www.atc.asia/events/SG2019_AGENDA.pdf
An investigation of the adoption of cloud computing by small to medium-sized enterprises (SMEs) in Nigeria	http://gsblibrary.uct.ac.za/researchreports/2017/Enumah.pdf
Islamic Crowdfunding: Fundamentals, Developments and Challenges	https://www.researchgate.net/profile/Muhammad_Musito/publication/331408070_Islamic_Crowdfunding_Fundamentals_Developments_and_Challenges/links/5c77f2a6a6fdcc4715a2bfce/Islamic-Crowdfunding-Fundamental
The Economics of Crowdfunding	https://link.springer.com/content/pdf/10.1007/978-3-319-66119-3.pdf
EMOTIONAL BANKING	https://link.springer.com/content/pdf/10.1007/978-3-319-75653-0.pdf
Digital banking manifesto: the end of banks?	http://ftp.shujuju.cn/platform/file/2018-01-09/4ccfbef11b284e2e953c4c5dc651bd22.pdf
Internationalization of SMEs: New Insights and Future Research Opportunities	https://ios.events.unibz.it/wp-content/uploads/2018/06/ISME-2018-Conference-Proceedings_def.pdf
FinTech in Sub-Saharan African Countries	http://www.africaniof.org/pdf/1561369197FTSSACEA.pdf
SHIRIN IMANBAEVA & UTZ DORNBERGER UNIVERSITÄT LEIPZIG	http://www.sept.uni-leipzig.de/fileadmin/user_upload/Publications/WP24_Imanbaeva_Dornberger_2017_Mapping_of_Innovative_Ways_to_Finance_SMES.pdf
Capital	https://www.lpea.lu/wp-content/uploads/2015/10/CAPITALV_06_final_pp.pdf
Blockchain & Transactions, Markets and Marketplaces	https://www.getsmarter.com/career-advice/wp-content/uploads/2017/07/mit_blockchain_transactions_report.pdf
El crecimiento de la industria fintech en China: un caso singular	http://www.iberchina.org/files/2019/fintech_china_bde.pdf
FinTech in Sub-Saharan African Countries: A Game Changer?	https://www.elibrary.imf.org/doc/IMF087/25592-9781484385661/25592-9781484385661/Other_formats/Source_PDF/25592-9781484399477.pdf
THE RISE OF TECHFINS: REGULATORY CHALLENGES	https://www.elgaronline.com/downloadpdf/edcoll/9781788979016/26_chapter13.xhtml.pdf
Identification of benchmarks for adoption by SMEs in the Indian cold chain	http://www.aims-international.org/aims15/15ACD/PDF/A209-Final.pdf
Investment funds and portfolio of loan guarantees as financial steps proposals to support innovative Small and Medium Enterprises.	http://store.ectap.ro/articole/1202.pdf
Fintechs and their emergence in banking services in CESEE	[illegible]

Figure 20: PDF files of “SME financing difficulties”

pdf-scholar_files

title	url
Financing difficulties of SMEs from its financing sources in China	https://file.scirp.org/pdf/JSSM_2014062510302476.pdf
Analysis on financing difficulties for SMEs due to asymmetric information	https://pdfs.semanticscholar.org/600e/c2e0e6df0fdaf3ecebb958259a08327886f6.pdf
Small business financing initiatives in Ghana	https://www.researchgate.net/profile/Nicholas_Biekpe/publication/265484299_Small_Business_Financing_Initiatives_in_Ghana_1/links/54e6f2990cf277664ff760a2/Small-Business-Financing-Initiatives-in-Ghana-1.pdf
SME development in China: A policy perspective on SME industrial clustering	http://www.eria.org/SME%20Development%20in%20China_A%20Policy%20Perspective%20on%20SME%20Industrial%20Clustering.pdf
Analysis on financing difficulties of small and medium-sized enterprises in China and corresponding countermeasures	https://pdfs.semanticscholar.org/42ac/7e7cc84ca991a7d605a8e569b03d5a140b06.pdf
WORKING CAPITAL FINANCING PREFERENCES: THE CASE OF MAURITIAN MANUFACTURING SMALL AND MEDIUM-SIZED ENTERPRISES (SMEs).	https://www.researchgate.net/profile/Kesseven_Padachi/publication/268685548_Working_capital_financing_preferences_The_case_of_Mauritian_manufacturing_small_and_medium-sized_enterprises_SMEs/links/547349b50
Financing canadian SME exporters	https://www.ic.gc.ca/eic/site/061.nsf/vwapj/Financing_Canadian_SME_Exporters_Eng.pdf/$FILE/Financing_Canadian_SME_Exporters_Eng.pdf
Four proposals for improved financing of SME development in ASEAN	http://www.academia.edu/download/33295754/finance_sme.pdf
Developing the capital market to widen and diversify SME financing: the Korean experience	https://www.asean.org/storage/images/archive/22633-8.pdf
Japanese banking regulations and SME finance under the global financial crisis	http://jsmeweb.org/jjmfe/pdf/jjmfe_2013august_4.pdf
Boosting SMEs with better and innovative financing services	http://www.aasmr.org/jsms/Vol4/No.2/JSMS_VOL4_NO2_002.pdf
Financing innovative small and mediumsized enterprises in times of crisis.	http://revecon.ro/articles/2011-2/2011-2-3.pdf
A Research of Financing Risk Management in Small and Medium-Sized Enterprises	https://www.scirp.org/pdf/18-1.8.3.pdf
SMEs Financing: the Extent of Need and the Responses of Different Credit Structures.	http://store.ectap.ro/articole/481.pdf
Perceptions of SME growth constraints in Nigeria	http://www.academia.edu/download/30383268/20110118110709280.pdf
The Access of Small and Medium Size Enterprises to Banking Financing and Current Challanges: The Case of EU Countries	http://oeconomica.uab.ro/upload/lucrari/1420122/21.pdf
The evaluation of the factors influence the access to debt financing by Tanzanian SMEs	https://pdfs.semanticscholar.org/5327/25b502bc3cebb89b0aba6b4623d92828afbe.pdf
A study on the influences of financing on technological innovation in small and medium-sized enterprises	https://pdfs.semanticscholar.org/e2ec/934c5cc9a6ad1e1d0d2c3e8482a97cfe5dc9.pdf
SME financing: Issues and Strategies	https://ishrathusain.iba.edu.pk/speeches/financialSector/SME_Financing_10_May_05.pdf
SME managers' financing preferences: The case of successful SMEs in Malaysia	https://core.ac.uk/download/pdf/20350078.pdf
Basel III and SME financing	https://library.fes.de/pdf-files/managerkreis/08528.pdf
Research on SME Financing and Its Development	https://webofproceedings.org/proceedings_series/ECOM/BUSEM%202018/BUSEM18051.pdf
The importance of leasing for SME financing	https://www.true-sale-international.de/fileadmin/tsi_downloads/TSI_kompakt/eif_euromoney_importance-of-leasing-for-smes_fv.pdf
Financing of small and medium enterprises (SMEs): Determinants of bank loan application	http://www.academicjournals.org/app/webroot/article/article1411119146_Abdesamed%20and%20Wahab.pdf
Involvement of accountants and bankers in SME-transfers: Perspectives and witnessed difficulties	https://www.researchgate.net/profile/Lex_Teeffelen/publication/307942228_Involvement_of_Accountants_and_Bankers_in_SME_Transfers_Perspectives_and_Witnessed_Difficulties/links/57d2cf4e08ae601b39a41d21/Involve
Capital structure and financing of SMEs: The Tunisian case	https://pdfs.semanticscholar.org/6d21/b6fa283b95726593d53d09004aef615e9225.pdf
Factoring–financing alternative for SMEs	http://mibes.teithessaly.gr/proceedings/2010/oral/14_final.pdf
Banking and SME financing in the United States	http://www.africres.org/SMME%20Research/SMME%20Research%20General/Conference%20Papers/Banking%20and%20SMME%20Financing%20in%20the%20US.pdf
Research on the Impact of Financial Ecological Environment on SME Financing.	http://www.ekolojidergisi.com/download/research-on-the-impact-of-financial-ecological-environment-on-sme-financing-5789.pdf
Difficulties in securing funding from banks: Success factors for small and medium enterprises (SMEs)	http://www.joams.com/uploadfile/2013/1024/20131024105843690.pdf
Access to Bank Credit and SME Financing	https://link.springer.com/content/pdf/10.1007/978-3-319-41363-1.pdf
SME Financing: Demand Side Problems and Supply Side Responses	http://ssadmin.bibm.org.bd/notice/02-07-19/Monograph%2003.pdf
Framework chapter: SME access to finance in selected East Asian economies	http://www.eria.org/uploads/media/Research-Project-Report/RPR_FY2010_14_Chapter_2.pdf
The challenges behind smes' access to debts financing in the Ghanaian financial market	https://pdfs.semanticscholar.org/500a/8f027ffb65eb8c3f8ab873453890f5753f15.pdf
Financing of small and medium enterprises (SMEs) in Libya: Determinants of accessing bank loan	https://www.idosijournals.org/mejsr/mejsr21(1)14/21.pdf
Considerations regarding growing financing capacity of SMEs	http://steconomiceuoradea.ro/anale/volume/2008/v3-finances-banks-accountancy/1001.pdf
Bank competition through the credit cycle: implications for SME financing	http://sme.ebi.gov.eg/Documents/Finance/Bank%20competition%20through%20the%20credit%20cycle%20implications%20for%20SME%20financing.pdf
Bridging the SME financing gap in Ghana: the role of microfinance institutions	http://www.academia.edu/download/35336924/OJBM_2014102911384020.pdf
SME Financing from Demand Side	https://www.researchgate.net/profile/Pengcheng_Luo/publication/273857713_China's_SME_Financing_Difficulty-_Study_from_Demand_Side/links/5511de050cf20bfdad5026cc.pdf
A research on impact of the capital structure and financing cost of small and medium sized enterprises (SME) on growth	http://www.academicjournals.org/app/webroot/article/article1380702799_Yurttadur%20and%20Kaya.pdf
SME Financing Constraints: Evidence from Survey of Construction Firms in Ghana	https://www.researchgate.net/profile/Alex_Eyiah/publication/326718582_SME_Financing_Constraints_Evidence_from_Survey_of_Construction_Firms_in_Ghana/links/5b606d880f7e9bc79a724ba8/SME-Financing-Constraints
SME development: narrowing the development gap in the ASEAN Economic Community	https://dirp3.pids.gov.ph/webportal/CDN/PUBLICATIONS/pidspjd12-sme%20development.pdf
Rural tourism development and financing in Romania: A supply-side analysis	https://www.agriculturejournals.cz/publicFiles/94_2014-AGRICECON.pdf
The impact of firm characteristics in access of financing by small and medium-sized enterprises in Tanzania	https://www.researchgate.net/profile/Alex_Kira/publication/272724027_The_Impact_of_Firm_Characteristics_in_Access_of_Financing_by_Small_and_Medium-sized_Enterprises_in_Tanzania/links/5891db48a6fdcc1b414684e
SME Financing of the Post-crisis Era in China Based on Blue Ocean Strategy	https://pdfs.semanticscholar.org/76a8/80f02f22738ad5be4c1da1183b99e2cfc3bd.pdf
SME Financing in Morocco: Issues and Alternatives	https://www.researchgate.net/profile/Abdelouhab_Salahddine/publication/319863853_SME_Financing_in_Morocco_Issues_and_Alternatives/links/5b9ed83f45851574f7d17c88/SME-Financing-in-Morocco-Issues-and-Alterna
Risk reduction in SME financing: Jeremie fund influence on financial situation of small and middle enterprises	http://www.academia.edu/download/376210/1ai7k00vw9rmre214l5d.pdf
SME strategic management and innovation—A comparative study between Finland and China	http://www.globelicsacademy.org/2011_pdf/Huang%20Zheng.pdf
SME financing in the Euro Area	http://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000444495/SME_financing_in_the_euro_area%3A_New_solutions_to_a.pdf
The Development of SME sector in Poland and New Instruments of Financing	https://www.researchgate.net/profile/Alina_Hyz/publication/263653520_The_Development_of_SME_sector_in_Poland_and_New_Instruments_of_Financing/links/0046353b6cdf44fa84000000.pdf
Small and medium enterprises' access to financing-A European concern: Evidence from Romanian SME	https://pdfs.semanticscholar.org/4bcc/f83008d4fc33c04a50504f4a824177dca23c.pdf
Financing constraints of SMEs in developing countries: Evidence, determinants and solutions	https://core.ac.uk/download/pdf/6635804.pdf
SME financing in Japan during the Global Financial Crisis: Evidence from firm surveys	http://hermes-ir.lib.hit-u.ac.jp/rs/bitstream/10086/26435/1/wp006.pdf
A study on financing of SMEs in Bangladesh	http://www.academia.edu/download/34227569/A_Study_on_Financing_of_SMEs_in_Bangladesh.pdf
Study on the Small and Middle Enterprises Financing Mode in Financial Crisis	https://pdfs.semanticscholar.org/9275/930da23b22d4048352e5978a5fe52acad285.pdf
Analysis of external financing use: A study of small and medium enterprises in Malawi	https://pdfs.semanticscholar.org/e59a/50e8b48abf73bc797200c3d5dfcc75d44bba.pdf
Alternative model for financing SMEs in Ghana	https://www.ijac.org.uk/images/frontImages/gallery/Vol._1_No._5/14.pdf
Small and medium enterprises financing in China	http://www.africres.org/SMME%20Research/SMME%20Research%20General/Conference%20Papers/SMME%20financing%20in%20China.pdf
„Study of Financing Preferences and Capital Structure among Malaysian SMEs: Evidence from Enterprise 50 Award Winners”	https://core.ac.uk/download/pdf/12008208.pdf
Financing sources for Romanian SMEs in the context of knowledge based economy	http://conference.management.ase.ro/archives/2015/pdf/64.pdf
BANK LOANS AND SMALL FIRM FINANCING IN ROMANIA.	http://www.oeconomica.uab.ro/upload/lucrari/1420121/15.pdf
Asymmetric Information and SME Financing: A case study of Homagama Divisional Secretariat.	https://www.researchgate.net/profile/Pasansa_Kodippiliarachchi2/publication/326185810_Asymmetric_Information_and_SME_Financing_A_case_study_of_Homagama_Divisional_Secretariat/links/5b3cea5ea6fdcc8506f503e
SME financing in the Netherlands	https://research.tue.nl/files/65722874/Master_Thesis_Casper_Miltenburg.pdf
Challenges experienced by young entrepreneurs entering the small and Medium Enterprise sector (SME) in Rajkot region, Gujarat-India	https://www.abacademies.org/articles/Challenges-experienced-by-young-entrepreneurs-entering-the-small-and-medium-enterprise-1528-2686-24-4-176.pdf
The impact of the global crisis on SME and entrepreneurship behavior–Romania and Finland cases	http://www.academia.edu/download/45561162/Article_914.pdf
SME Credit Guarantee in China and Risk Evaluation	https://cpb-us-w2.wpmucdn.com/sites.wustl.edu/dist/0/590/files/2016/12/Yvonne_SME-credit-rating-2jjto7b.pdf
The role of banks in small and medium enterprises financing: a case study from Kosovo	https://www.researchgate.net/profile/Florin_Peci/publication/272805936_The_Role_of_Banks_in_Small_and_Medium_Enterprises_Financing_A_Case_Study_from_Kosovo/links/5969df4d0f7e9b8091943465/The-Role-of-Banl

Figure 21: HTML files of “SME financing difficulties”

web-scholar_files

title	url
SME financing in emerging markets: Firm characteristics, banking structure and institutions	https://www.tandfonline.com/doi/abs/10.2753/REE1540-496X500107
SME financing in the UK and in China: a comparative perspective	https://www.emerald.com/insight/content/doi/10.1108/14626000610705769/full/html?queryID=51%2F5408313
SME firm performance-financial innovation and challenges	https://www.sciencedirect.com/science/article/pii/S1877042815038409
Bank market structure, competition, and SME financing relationships in European regions	https://link.springer.com/article/10.1007/s10693-009-0060-0
The factors of affecting financing and countermeasures based on asymmetric information and marginal information cost	https://link.springer.com/chapter/10.1007/978-3-642-29084-8_104
Countermeasures for Rural Financing Difficulties in Western China under Double Dual Structure——Research Based on Evolutionary Game Theory [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-DJKX201101003.htm
Banking system soundness is the key to more SME financing	https://www.econstor.eu/handle/10419/106312
A Study of the Connecting Channel between SME Financing and Private Funds: Analysis and Reflection on the Fund Management System of Informal Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JXCZ201101007.htm
Economic Analysis of the Financing Difficulties for Private SME [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CQYD200906009.htm
Group Financing Model and Path Direction——Based on Game Analysis of the SME Decision-making and Rational Breakout [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CJKX200905007.htm
Financing System Study of SME under New Institutional Economics [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-XTGS201106020.htm
The financing of technology-based small firms: A review of the literature	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=759544
Economic Analysis on Financing Credit Chain of Medium and Small Enterprises [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-HNJR200706010.htm
The Routing Selection of the SME's Financing Platform Construction [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-HLJC201004005.htm
Blockchain, bank credit and SME financing	https://link.springer.com/article/10.1007/s11135-018-0806-6
Loan rate pricing of SME financing based on agent-based computational finance approach	https://www.sciencedirect.com/science/article/pii/S1874865110600877
Different Guarantee Institutions and Their Effects on the Financing of Small and Medium-Sized Enterprises (SMEs)[J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JJYJ200509006.htm
Interest Rate Marketization, Relationship Lending and SME Financing	http://en.cnki.com.cn/Article_en/CJFDTotal-XCXB201404006.htm
Impediments to SME access to finance and credit guarantee schemes in China	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=486204
Fisheries SMEs Financing Difficulties and Countermeasures	http://en.cnki.com.cn/Article_en/CJFDTotal-NYJG201206019.htm
The Analysis of civil Financing theory: the scope, the Comparison and the InstitutionalTransition [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JRYJ200408012.htm
Analysis on SME Financing Difficulty, Private Lending Difficulty and Institutional Reform [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-DJGL201302016.htm
Financing the growth of SMEs in Africa: What are the contraints to SME financing within ECOWAS?	https://www.sciencedirect.com/science/article/pii/S1879933717300362
Comparison analysis of small and medium enterprise financing structure—— questionnaire basing on Jiangshu province's small and medium enterprises [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TAJJ200803015.htm
Research on SME Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-SGXB200905011.htm
Financing small businesses in France	https://www.econstor.eu/handle/10419/44830
Causes for SME's Financing Problem in China and its Countermeasures [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-YSXS200403020.htm
Financing high-tech SMEs in China: A three-stage model of business development	https://www.tandfonline.com/doi/abs/10.1080/08985620903233937
A Study on the Structure of Financing Demand of Small and Medium Enterprises and the Counter-measures [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CSJR200908011.htm
An Analysis of perfection of SME Legal System in China [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TYDX200703015.htm
Changes in Interest Rate, Shadow Banking and SME Financing	http://en.cnki.com.cn/Article_en/CJFDTotal-CSJR201307007.htm
An empirical evidence of small business financing in China	https://www.emeraldinsight.com/doi/abs/10.1108/01409170810920666
On The Financing Predicament and Strategy of SME [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JGGB200902012.htm
The Research on Inventory Pledge in SME Financing	http://search.proquest.com/openview/3efdb691503fd69572deb9f49b8fdb82/1?pq-origsite=gscholar&cbl=2028701
Relational Financing in Small and Medium-sized Enterprises: empirical study and interpretation [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-SXCJ200402021.htm
Bank competition, government intervention and SME debt financing	https://www.emeraldinsight.com/doi/abs/10.1108/CFRI-02-2017-0007
Information, Informal Finance and SME Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JJYJ200507003.htm
Liaonmg small and middle enterprise financing difficulty forming reason and countermeasure research [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TAJJ201209065.htm
An Analysis of Financing Difficulty of Small and Medium Enterprises in the New Rra [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-ZGSG201205012.htm
On Legislation of Private Financing [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-NFZJ201201010.htm
Research on Financing Cost of Small and Medium-Sized Enterprises by Internet Finance	https://www.scirp.org/html/7-1761631_80477.htm
Relationship lending and SME financing in the continental European bank-based system	https://link.springer.com/article/10.1007/s11187-008-9129-7
Analysis on financing problems for small and medium-sized enterprises	https://ieeexplore.ieee.org/abstract/document/6013702/
The evolutionary game analysis of credit behavior of SME in guaranteed loans organization	https://www.sciencedirect.com/science/article/pii/S1877050913002512
Intellectual property impawn solving science style small and medium enterprise financing problem [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-TAJJ200902089.htm
Silicon Valley Bank Mode and Its Implication on Financing of S&T SMEs in China [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CHXI201005018.htm
Empirical Study of Factors Affecting the Financing Ability of Small-and Medium-Sized Real Estate Enterprises	https://ascelibrary.org/doi/abs/10.1061/9780784413135.078
Research on SME′ s Financing Problem in China: in View of Government Behavior [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-KJRC200802008.htm
The Research on the Financing of Small and Medium-Sized Enterprises——Based on the Financing Innovative Angle [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JZGC200701017.htm
Mezzanine financing-a feasible option of SME) s financing	http://en.cnki.com.cn/Article_en/CJFDTotal-ZGKT201103010.htm
Access to financing in a small island economy: Evidence from Mauritius	https://www.tandfonline.com/doi/abs/10.1080/15228916.2012.727753
Reasons and Countermeasures of Difficulties in Financing of Small And Medium-sized Enterprises'	http://en.cnki.com.cn/Article_en/CJFDTotal-MGZB200901024.htm
Small and medium business management-financial sources and difficulties	https://www.learntechlib.org/p/176804/
The research of external financing options of small and medium-sized enterprises in China	http://www.cscanada.org/index.php/mse/article/view/j.mse.1913035X20120603.Z0279
Bank deleveraging, the move from bank to market-based financing, and SME financing	https://www.oecd-ilibrary.org/finance-and-investment/bank-deleveraging-the-move-from-bank-to-market-based-financing-and-sme-financing_fmt-2012-5k91hbvfh9g3?crawler=true
The Role Study about the Alliances Between Banks and Enterprises in SMEs Financing Guarantee [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JXJR201103013.htm
Innovative approach to SME financing in Nigeria: A review of small and medium industries equity investment scheme (SMIEIS)	https://www.tandfonline.com/doi/abs/10.1080/09718923.2005.11893635
Research on the Constraints and Countermeasure of SME Clusters Financing——Taking Fujian Province as the Example	http://en.cnki.com.cn/Article_en/CJFDTotal-MGXY201103008.htm
SME Financing Difficulties and the Strategies	http://en.cnki.com.cn/Article_en/CJFDTotal-JJSY201322031.htm
Cargo Backed Loans for SMEs to Overcome Their Financing Difficulties [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-CLSJ200801009.htm
An analysis of causes for SMEs financing difficulty	http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.653.9072&rep=rep1&type=pdf
Research on the financing of small and medium enterprises	https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.656.2748&rep=rep1&type=pdf
Islamic financing in mitigating access to financing problems of SMEs in Malaysia: a survey analysis	https://journals.iium.edu.my/intdiscourse/index.php/islam/article/view/926
On China's Private Lending Crisis and Preventive Measures Recommended——From the Perspective of Private Financing for SMEs [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-HNJG201307017.htm
What are the extent of Small and Medium-sized Enterprises Financing Problems in China and Its Countermeasures: Based on SME financing system and cases of ..	http://www.diva-portal.org/smash/record.jsf?pid=diva2:488199
Financing small and medium-scale contractors in developing countries: a Ghana case study	https://www.tandfonline.com/doi/abs/10.1080/0144619032000111241
The causes and countermeasures of the difficulties in China's small business financing——Analysis based on province questionnaire [J]	http://en.cnki.com.cn/Article_en/CJFDTotal-JJXJ201008014.htm

Figure 22: HTML files of "Fintech SME solutions"

web-scholar_files_solution

title	url
Fintech solutions for small businesses	https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119218906.ch32
The Fintech Book: the Financial Technology Handbook for Investors, Entrepreneurs and Visionaries	https://journal.ugm.ac.id/jieb/article/download/23554/15535
The promise of fintech in emerging markets: Not as disruptive	https://go.gale.com/ps/i.do?id=GALE%7CA570046846&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=20840845&p=AONE&sw=w
Beyond Fintech	https://openknowledge.worldbank.org/handle/10986/30371
The Utilization of FinTech (P2P Landing) as SME's Capital Solution in Indonesia: Perspective in Islamic Economics (Qirad)	http://www.ejournal.iainpurwokerto.ac.id/index.php/icms/article/download/1818/1283
Analyzing China's fintech industry from the perspective of actor–network theory	https://www.sciencedirect.com/science/article/pii/S0308596115001810
FINTECH in Scotland: building a digital future for the financial sector	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2839696
FinTech and SMEs: The Italian Case	https://www.igi-global.com/chapter/fintech-and-smes/221589
Is Fintech a Solution for Financial Inclusion and Women Empowerment?	http://scholar.dkyobobook.co.kr/searchDetail.laf?barcode=4010027326443
SMEs-Fintech: Support Increase of BMT through financing and recruitment member based of technology	http://ejournal.upi.edu/index.php/rief/article/view/13858
Fintech in a Flash: Financial Technology Made Easy	https://books.google.com/books?hl=en&lr=&id=smSEDwAAQBAJ&oi=fnd&pg=PT17&dq=Fintech+SME+solutions&ots=4lx8En6bpu&sig=XYPGAyoQiQRUVtYRyczTfqMS_Xc
Mittelstandsfinanzierung Im Kontext Von FinTech Und Digital Finance (SME Financing in the Context of FinTech and Digital Finance)	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3083908
Fintech, Inclusive Growth and Cyber Risks: Focus on the MENAP and CCA Regions	https://books.google.com/books?hl=en&lr=&id=rgZxDwAAQBAJ&oi=fnd&pg=PP1&dq=Fintech+SME+solutions&ots=g4YHVlYpS1&sig=qa0Q5ds7pQEEpfr4IFTKt8GIH0Y
FinTech (financial technology): what is it and how to use technologies to create business value in FinTech way?	http://glyndwr.repository.guildhe.ac.uk/id/eprint/17310
The habitual selection of SME funding in an increasingly diverse market	https://www.narcis.nl/publication/RecordID/oai:hbokennisbank.nl:sharekit_hu%3Aoai%3Asurfsharekit.nl%3A85d9f0e5-b08e-4076-8cf7-9cc1208761c5
Addressing the SME finance problem	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3249560
From FinTech to TechFin: the regulatory challenges of data-driven finance	https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nyujolbu14&section=14
Fintech startup: analysis of a digital lending business for small & medium-sized enterprises (SMEs) in Argentina	https://www.politesi.polimi.it/handle/10589/139356
Handbook of blockchain, digital finance, and inclusion: Cryptocurrency, FinTech, InsurTech, regulation, ChinaTech, mobile security, and distributed ledger	https://books.google.com/books?hl=en&lr=&id=4VEADQAAQBAJ&oi=fnd&pg=PT20&dq=Fintech+SME+solutions&ots=w9VFG9GkmJ&sig=lpKVNbsb6yD7UygnrovDrCEXtg4
The Fintech Revolution: A Threat to Global Banking?	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3255725
SME Finance in Asia: Recent Innovations in Fintech Credit, Trade Finance, and Beyond	https://www.adb.org/publications/sme-finance-asia-innovations-fintech-credit-trade-finance-beyond
SME funding remains elusive	https://journals.co.za/content/journal/10520/EJC-16a4c799bf?crawler=true&mimetype=application/pdf
FinTech in Taiwan: a case study of a Bank's strategic planning for an investment in a FinTech company	https://www.biomedcentral.com/openurl?doi=10.1186/s40854-016-0037-6
Go-Mudaraba: The Solution of Poverty and Unemployment in the Digital Era	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215243
The Role of Fintech Encourages the Export of Small Medium Enterprises in Indonesia	https://ojs.amhinternational.com/index.php/jsds/article/view/2478
The Fintech Market in Germany	https://link.springer.com/chapter/10.1007/978-3-319-54666-7_4
Financing Models for SMEs in the Age of Disintermediation	https://link.springer.com/chapter/10.1007/978-3-030-02831-2_5
Fintech: Ecosystem, business models, investment decisions, and challenges	https://www.sciencedirect.com/science/article/pii/S0007681317301246
Exploring fintech solutions for women	https://idl-bnc-idrc.dspacedirect.org/handle/10625/57158
Fintech and the financing of SMEs and entrepreneurs: From crowdfunding to marketplace lending	https://link.springer.com/chapter/10.1007/978-3-319-66119-3_6
Fintech and the Transformation of the Islamic Finance Regulatory Framework in Malaysia	https://www.emerald.com/insight/content/doi/10.1108/978-1-78973-545-120191018/full/EthisKapital.com
Embracing the promise of fintech	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3419994
Development of Islamic FinTech: Transforming the Islamic Banking and Finance Regulatory Framework in Malaysia.	https://www.researchgate.net/profile/Azwina_Wati_Manaf/publication/335029438_Development_of_Islamic_FinTech_Transforming_the_Islamic_Banking_and_Finance_Regulatory_Framework_in_Malaysia/links/5d4bbd1e458
FINTECH IN ISLAMIC FINANCE	https://books.google.com/books?hl=en&lr=&id=YAWcDwAAQBAJ&oi=fnd&pg=PT57&dq=Fintech+SME+solutions&ots=Vh52gLRJIL&sig=e7R1b5TqsC1r5bqXhtKFayU10_w
FinTech in the Saudi Context: Implications for the Industry and Skills Development	https://www.igi-global.com/chapter/fintech-in-the-saudi-context/237267
A Seat at the Table–Bringing the Voice of FinTech to the US Regulatory Process	https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119362197.ch18
Finance strategies for medium-sized enterprises: FinTech as the game changer	https://www.igi-global.com/chapter/finance-strategies-for-medium-sized-enterprises/207822
Risk-sharing securities: Accelerating finance for SMEs	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3150197
BISAIN APP AS A SOLUTION OF SHARIA FUNDING FOR SMES IN INTEGRATED SYSTEM	https://grdspublishing.org/index.php/people/article/view/1748
Alternative financing sources supporting SMEs growth	http://eprints.ugd.edu.mk/20967/
Fintech and the financing of entrepreneurs: From crowdfunding to marketplace lending	https://papers.ssrn.com/Sol3/papers.cfm?abstract_id=2967891
The evolution of Fintech: A new post-crisis paradigm	https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/geojintl47&section=41
SME Microfinance, Fractional Ownership and Crowdfunding	https://link.springer.com/chapter/10.1007/978-3-319-76014-8_12
Marzunisham Omar: Fintech & entrepreneurship	https://www.bis.org/review/r171017a.htm
Sustainability, FinTech and Financial Inclusion	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3478663
Women and trade: Gender's impact on trade finance and fintech	https://digitalcommons.ilr.cornell.edu/intl/680/
Fintech Hypes, but Wealthy Internet Savvy Investors Prefer to Stay Hybrid	https://link.springer.com/chapter/10.1007/978-3-662-49275-8_33
The development of collaborative model between fintech and bank in Indonesia	https://dl.acm.org/citation.cfm?id=3234783
Fintech Emergence and Global Evolution	https://link.springer.com/chapter/10.1007/978-3-030-24666-2_2
21 Geographic decentralization of FinTech companies in Sweden	https://books.google.com/books?hl=en&lr=&id=2T1MDwAAQBAJ&oi=fnd&pg=PT418&dq=Fintech+SME+solutions&ots=ubn0a8p3CM&sig=giyqUFbY1DpNLDWWfdB4HLfCJLo
Fintech Explained	https://books.google.com/books?hl=en&lr=&id=xshIDwAAQBAJ&oi=fnd&pg=PT201&dq=Fintech+SME+solutions&ots=plqha73PmZ&sig=fV3dPwPlSBXzyNxvdoij3iWz9HQ
Evaluating the financing needs of construction sector SMEs in the Colombian context	http://openresearch.ocadu.ca/id/eprint/2434/
A REVIEW OF FINTECH REGULATIONS IN EMERGING COUNTRIES'ECONOMIES–CHINA, SINGAPORE AND HONG KONG	https://www.atlantis-press.com/proceedings/iclgg-17/25902323
Study of Customer Preferences Towards Fintech Firms in Mumbai Suburbs	http://search.proquest.com/openview/d8c5ec32c04c73d8468806e933afbd5d/1?pq-origsite=gscholar&cbl=2046061
Predictive Algorithms–Building Innovative Online Banking Solutions	https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119218906.ch25
FinTech 2.0: Software as the future of payments distribution	https://www.ingentaconnect.com/content/hsp/jpss/2019/00000013/00000003/art00006
FINANCIAL INTERMEDIATION, FINTECH AND SHARĪ 'AH COMPLIANCE	https://books.google.com/books?hl=en&lr=&id=YAWcDwAAQBAJ&oi=fnd&pg=PT71&dq=Fintech+SME+solutions&ots=Vh52gLRJJD&sig=HMgWbG1wGUz2T3zodbNlrZ90UBM
Fintech: The New DNA of Financial Services	https://books.google.com/books?hl=en&lr=&id=UBiBDwAAQBAJ&oi=fnd&pg=PR7&dq=Fintech+SME+solutions&ots=zfm81oNQMq&sig=epSUtcrqfp2J228lw0kIO-EatL4
Preface: FinTech as a Disruptive Technology for Financial Institutions	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3376358
Forecasts for the FinTech Market in Germany	https://link.springer.com/chapter/10.1007/978-3-319-54666-7_6
Fintech in Islamic Finance: Theory and Practice	https://books.google.com/books?hl=en&lr=&id=sZGzDwAAQBAJ&oi=fnd&pg=PR7&dq=Fintech+SME+solutions&ots=-NPdQ-Ojcc&sig=XyTu5_QhcFQZ_KApzfEeXl1l_tU
ISLAMIC FINANCIAL TECHNOLOGY (FINTECH): ITS CHALLENGES AND PROSPECT	https://www.atlantis-press.com/proceedings/assdg-18/55912844
Fintech Regulation	https://link.springer.com/chapter/10.1007/978-3-030-24666-2_8
Fintech as a Source of Financial Innovations on the Polish Financial Services Market	https://ageconsearch.umn.edu/record/288528/
FinTech Activities and Business Models: Analogies and Differences with the Traditional Financial Channels	https://link.springer.com/chapter/10.1007/978-3-030-22426-4_2
Regulating FinTech Businesses: The Malaysian Experience	https://www.igi-global.com/chapter/regulating-fintech-businesses/236803
Promoting SME Finance in the Context of the Fintech Revolution: A Case Study of the UK's Practice and Regulation	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144767

Under the search of 'SME financing difficulties,' there are a total of 396 HTML files and 80 PDF files. Under the search of 'Fintech SME solutions', there are a total of 541 web files and 151 PDF files. Because we identify the problem in figure 17, we will only keep PDF files and continue our research.

3.2 Text Mining

3.2.1 Corpus

Before doing text clustering, we must first set up a predictive database to uniformly process the text, and it is a corpus. The document corpus contains the following features:

- The number of documents: There are crawled 1,000 articles from the Internet, but filtered to leave only PDF files as the content of the corpus. The total number of PDF files is 231.

- Documents: Because articles scraped from the Internet are formatted as PDF files. We must convert PDF files from PDF to txt format.

- High dimensionality: If a unique term is treated as one dimension, 231 will exceed 500 unique terms. So the term-document matrix for the corpus will be 231 * 500. This high dimensionality will significantly increase memory and time complexity in the next clustering algorithms.

- Noisy data: When dealing with noisy data, noisy data should be determined based on the situation. We will discuss about the part in the following steps.

After converting PDF format to txt format, we will have a corpus of 172 text files. Because some PDF files are either encrypted or "broken".

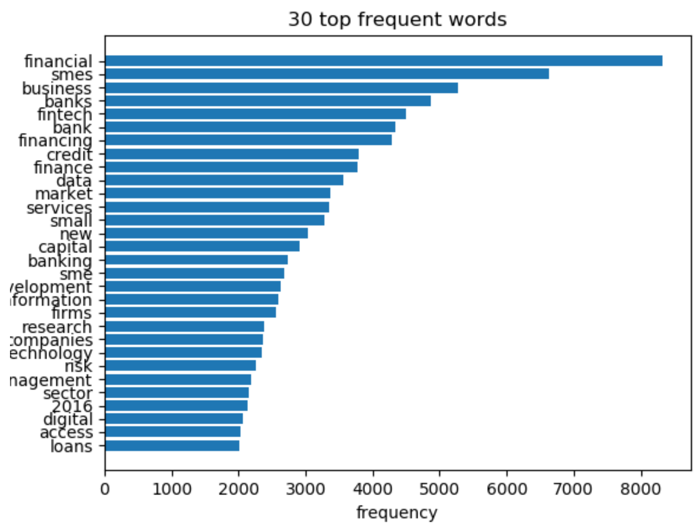
3.2.2 Preprocessing

It is crucial to remove the list of stopwords, and stopwords removal is the process of removing non-informative bearing words to reduce noise.
After removing stopwords, we have a comprehensive understanding of the dataset. We visualize the top 30 most frequent words. In figure 23, we can see that some words of similar or the same meanings appear multiple times, and they are only in different morphology. For example, both "financial" and "financing" appear in the top 30 frequency

word. However, for us, they convey the same meaning, and it is necessary to convert them in the same form.

Moreover, we can see that the number "2016" is shown in the most frequent list. These numbers should be removed as well. When we enlarge our most frequent words list, we can find more problems. For example, we show the top 50 frequent words in figure 24, and we can see that "et" is not an English word. We need to remove these non-English words as well.

Figure 23: Visualization for top 30 words



To summarize, we need several more steps to remove noisy data in our dataset. First, we need to convert our alphabets to lower cases.

Second, functional words are prevalent. Compared with other words, functional words have no practical meaning, such as "the," "is," "at," "which," "on" and etc. These stop-words do not contribute much to the analysis of the articles, and we remove them from the text during the preprocessing to better capture text features and save space.

Third, we need to remove punctuation and numbers. Forth, we have to remove short words, and we define short words as words with two letters or less. Fifth, we need to detect the language of strings, and in our research, only the articles in English will be kept. We mainly rely on the "langdetect" function to determine the language of a set of

Figure 24: The most frequent 50 words

```
('financial', 8333)
('smes', 6626)
('business', 5270)
('banks', 4866)
('fintech', 4497)
('bank', 4342)
('financing', 4294)
('credit', 3784)
('finance', 3779)
('data', 3571)
('market', 3367)
('services', 3351)
('small', 3273)
('new', 3031)
('capital', 2918)
('banking', 2737)
('sme', 2687)
('development', 2637)
('information', 2592)
('firms', 2565)
('research', 2388)
('companies', 2364)
('technology', 2343)
('risk', 2255)
('management', 2193)
('sector', 2144)
('2016', 2132)
('digital', 2057)
('access', 2036)
('loans', 2011)
('countries', 2000)
('2017', 1935)
('innovation', 1881)
('industry', 1836)
('growth', 1793)
('lending', 1783)
('economic', 1774)
('international', 1717)
('loan', 1698)
('enterprises', 1697)
('use', 1686)
('et', 1646)
('2018', 1625)
('firm', 1575)
('investment', 1549)
('institutions', 1538)
('10', 1497)
('government', 1473)
('journal', 1465)
('equity', 1412)
```

strings. We will use detect_langs, which returns a list of language with probabilities, and then iterate through this list, returning the language if the option is English. Or false if this is not the case. This function works well for this purpose.

Sixth, we need to perform lemmatization and stemming process. We will do them separately. Before we do them, we need to know part-speech tagging. According to its definition (Sachin and Divya 2018), part-of-speech tagging (pos tagging) is the process of marking up a word in a corpus as corresponding to a particular part of speech, based on both its definition and its context. For lemmatization, we need to perform parts of speech tagging on the text. This processing is to facilitate syntactic and semantic analysis. For texts with a small amount of content, in nltk, it is easy to process by nltk.pos_tag (). This method only works for lists of words. But for large texts, it is difficult to make the pos tag look coherent like texts with a few sentences. Therefore, we will find coherent pos tag for each word and map it to the right input character that the WordnetLemmatizer accepts and passes it as the second argument to lemmatize(). nltk.pos_tag() returns a tuple with the pos tag. We map nltk's pos tags to the format wordnet lemmatizer would accept. The get_wordnet_pos() function defined does mapping job. We first initiate a lemmatizer. Then we lemmatize them with the appropriate pos tag on our text. For stemming, we will use the most classic porterstemmer to stem the text. At the same time, stemming applies algorithmic rules to generate stems, and it uses the rules to decide whether it should strip a suffix or not.

3.2.3 Directory reset

After implementing these preprocessing steps, the directory path of the corpus after the preprocessing part is set and then stored to two specified paths for lemmatization and stemming.

We will add pre_dataset_lemmatizing corpus based on the path specified by jupypterlab. Then we unify the file paths via os.path.join. Then we use the function os.makedirs (path) to create multiple levels of directories at once. At the same way, we create a corpus for pre_dataset_stemming.

3.2.4 The most frequent words removal

In each article, the most frequent words are filtered based on statistical calculations, but this method can lead to two potential problems.

First, the most frequent words filtered from each article may be different. The high-frequency words filtered by one article may not be high-frequency words in another article. So all documents need to be integrated to finally filtered high-frequency words, and TF-IDF will be used when dealing with frequency calculations.

Second, from figure 23 and 24, we can see that not all high-frequency words have research value. For example, "bank," "business," and other words will repeatedly appear in each document, so these words must be eliminated. The total number of words is 34930. When we visualize the frequency of words, we find that around the top 50 words are non-descriptive words. So we decide to remove the top 50 words to continue our research.

In the processing of word frequency, we first tokenize the words, and then we will use "countvectorizer" function, which converts the words in the text into a word frequency matrix together with the "fit_transform" function.

The following are the words that were removed. In figure 25, we can the words removed based on the difference of lemmatization and stemming.

Figure 25: The most frequent 50 words removal

```
remove_50_top_words_from_corpus("dataset_lemmatizing")
remove_50_top_words_from_corpus("dataset_stemming")
remove_50_top_words_from_corpus("dataset_full")
```

```
['bank', 'financial', 'business', 'smes', 'market', 'fintech', 'service', 'financing', 'credit', 'use', 'firm', 'data', 'company', 'fina
nce', 'technology', 'loan', 'small', 'new', 'risk', 'sme', 'capital', 'country', 'banking', 'development', 'access', 'provide', 'informa
tion', 'innovation', 'sector', 'research', 'payment', 'investment', 'management', 'enterprise', 'industry', 'digital', 'customer', 'mode
l', 'http', 'study', 'high', 'fund', 'growth', 'need', 'institution', 'make', 'lending', 'economic', 'process', 'policy']
['bank', 'sme', 'financi', 'financ', 'busi', 'market', 'fintech', 'servic', 'develop', 'use', 'credit', 'technolog', 'firm', 'data', 'co
mpani', 'innov', 'loan', 'provid', 'new', 'manag', 'risk', 'fund', 'capit', 'inform', 'small', 'countri', 'invest', 'access', 'researc
h', 'sector', 'institut', 'digit', 'industri', 'econom', 'payment', 'custom', 'enterpris', 'intern', 'model', 'http', 'studi', 'govern',
'lend', 'product', 'process', 'growth', 'need', 'rate', 'includ', 'differ']
['bank', 'sme', 'financi', 'financ', 'busi', 'market', 'fintech', 'servic', 'develop', 'use', 'credit', 'technolog', 'firm', 'data', 'co
mpani', 'innov', 'loan', 'provid', 'small', 'new', 'manag', 'risk', 'fund', 'capit', 'inform', 'countri', 'invest', 'access', 'researc
h', 'sector', 'institut', 'digit', 'industri', 'econom', 'payment', 'custom', 'enterpris', 'intern', 'model', 'http', 'studi', 'high',
'govern', 'lend', 'product', 'process', 'growth', 'need', 'make', 'rate']
```

After removing the most frequent 50 words, we revisualize the top 30 frequent words. We can see the result of visualization after stemming and lemmatization in figure 26 and figure 27.

Figure 26: The visualization for word frequency count after stemming

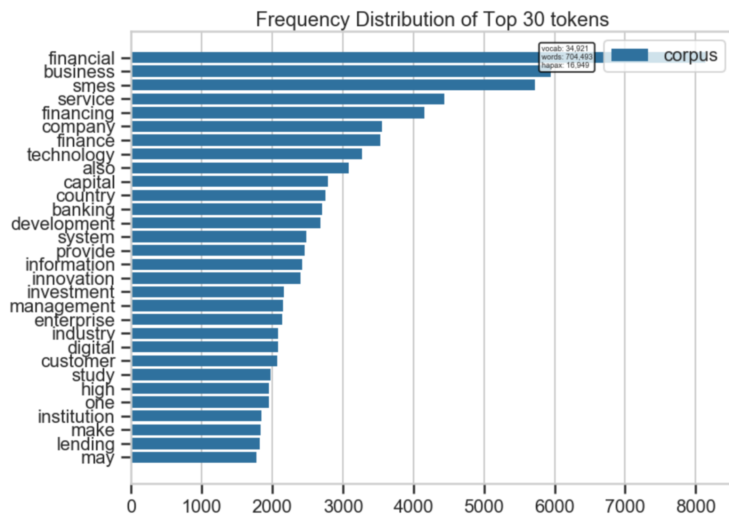
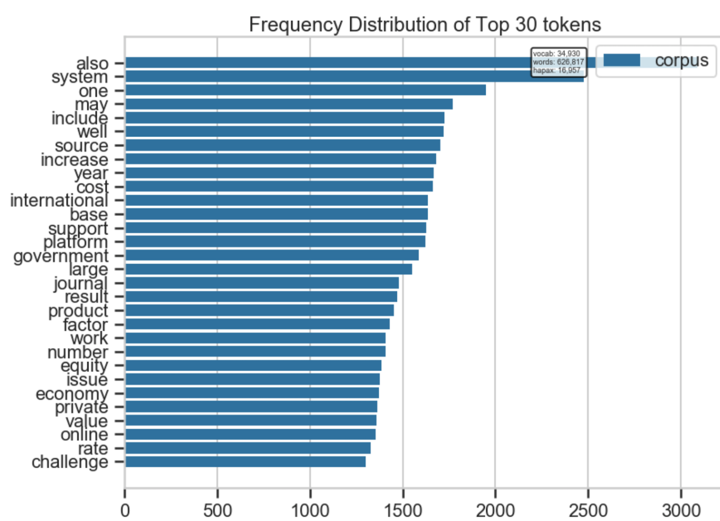


Figure 27: The visualization for word frequency count after lemmatization

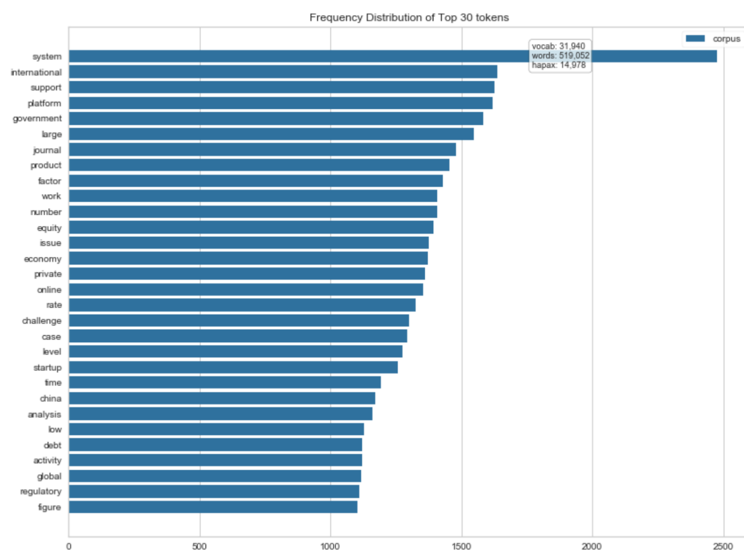


Adjustment for preprocessing

When we observe that the frequent words in figure 27, we can see that words such as "also" are not removed. After reviewing these articles, we need to add more stopwords, and we used a list stopwords of ranks.nl.

Besides, we also need to normalize a Unicode string. We use unicodedata.normalize to normal form KD (NFKD), and it will apply the compatibility decomposition, i.e. replace all compatibility characters with their equivalents. After adjustment of the problems, we revisualize our top 30 frequent tokens in figure 28.

Figure 28: The re-visualization for word frequency count after lemmatization



It can be seen from the graph that "stemming" is not as thorough as "lemmatization" in dealing with word frequency problems. In our next research, we will conduct our research only based on the result of lemmatization.

3.3 Analysis

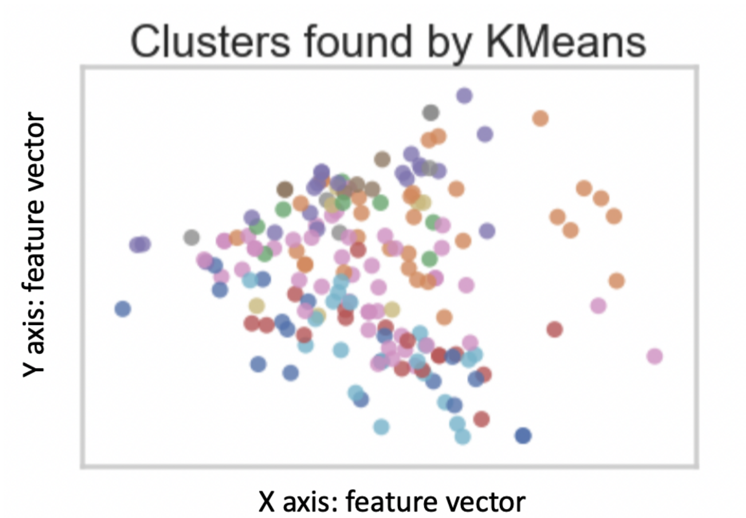
3.3.1 Comparing clustering algorithms

From the literature review, we know that many clustering algorithms can be used. Now, we first select suitable clustering algorithms. We will visualize K-means/agglomerative clustering/DBSACN and see if they can deliver good results.

Our dataset is the list of multiple dimensional vectors (168*168 by applying SVD), so in order to show our dataset in two-dimensional space. We only use the first two feature vectors. Both the x/y axis are feature vectors. The dots of the same colour represent that they are in the same cluster.

From figure 29 to figure 31, we can see the result of visualization for each clustering algorithms.

Figure 29: The visualization for K-means



From figure 31, we can see that DBSACN has many black dots, which represent noise data. Therefore, we will use K-means and agglomerative clustering for our research.

Figure 30: The visualization for Agglomerative clustering

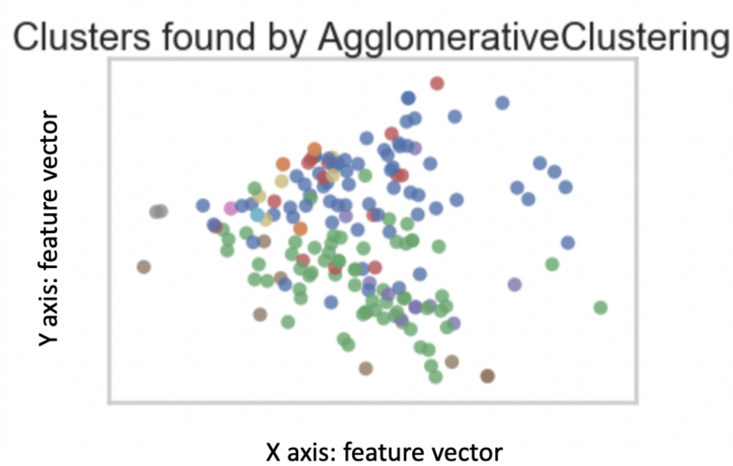


Figure 31: The visualization for DBSACN

3.3.2 Choosing the best k

The next step is to decide the best k for our dataset, and we will use elbow method to determine the best value of k. *Elbow method is a method which looks at the percentage of variance explained as a function of the number of clusters. This method exists upon the idea that one should choose a number of clusters so that adding another cluster doesn't give much better modelling of the data. The percentage of variance explained by the clusters is plotted against the number of clusters* (Bholowalia and Kumar 2014).

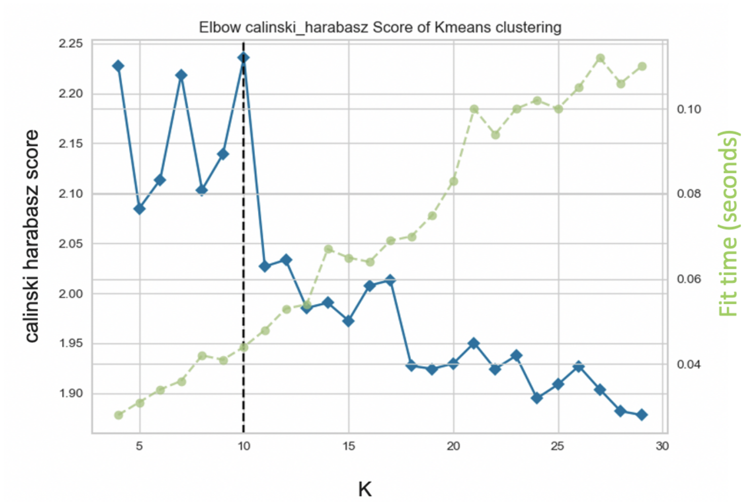
When these overall metrics for each model are plotted, it is possible to visually determine the best value for k. If the line chart looks like an arm, then the "elbow" (the point of inflection on the curve) is the best value of k. The "arm" can be either up or down, but if there is a strong inflection point, it is a good indication that the underlying model fits best at that point (scikit-yb developers 2019). To choose the best k value, we need to calculate tf-idf.

We will use Latent Semantic Analysis (LSA). *Latent Semantic Analysis is a corpus-based statistical method for inducing and representing aspects of the meanings of words and passages (of natural language) reflective in their usage* (Maletic and Valluri 1999).

LSA relies on a Single Value Decomposition (SVD) of a matrix (word context) derived from a corpus of natural text that pertains to knowledge in the particular domain of interest. SVD is a form of factor analysis and acts as a method for reducing the dimensionality of a feature space without serious loss of specificity. Typically the word by context matrix is very large and (quite often) sparse. SVD reduces the number of dimensions without great loss of descriptiveness. Single value decomposition is the underlying operation in a number of applications including statistical principal component analysis, text retrieval, pattern recognition and dimensionality reduction, and natural language understanding (Maletic and Valluri 1999). LSA will be used in the context of k-means clustering.
After dimension reduction, we can conduct the elbow mehtod. In figure 32, the x-axis shows the value of k and y-axis shows an average score for all clusters. In our research, we will use the calinski harabasz score, which computes the ratio of dispersion between

and within clusters.

Figure 32: calinski_harabasz



Because by default, we set "locate_elbow" to true. Therefore, the black line indicates the result of the k value automatically filtered by the system.

The green dashed line is used to display the amount of time to train the clustering model per as a function of k.

The best k is ten from the graph, and we will continue our research based on this result.

3.3.3 K-means when k=10

To perform dimensionality reduction, we will use Singular Value Decomposition (SVD) again, *In order to improve the traditional K-means algorithm in the clustering process, it is difficult to accurately preset the number of k clusters. The clustering results are affected by the initial center, sensitive to noise points, unstablility and other shortcomings. At the same time, for the problems of high data dimensionality, sparse spatial distribution, and latent semantic structure in text clustering, singular value decomposition is added (SVD) for rough classification, combined with K-means algorithm's English text clustering optimization algorithm (SVD-K-means). The basic logic is based on the SVD method to get the subject of the text* (Dai Yueming et al. 2018).

Before formal K-means, we will use "tfidfvectoriozer "and "fit_transform" function to count the number of times each word in the dataset occurs and project that count into a vector (tf-idf vector). Then, to have K-means behave as spherical k-means for better results, LSA / SVD results will be normalized. We sort cluster centers by proximity to centroids.

Then through the classification of the article, we label the label class for each cluster. We also count the articles in each cluster. And we will get the result in figure 33.

Figure 33: Label class for each cluster (K-means)

```
Cluster 0:['cid', 'lease', 'debt', 'variable', 'china', 'guarantee', 'journal', 'cost', 'rate', 'difficulty']
Cluster 0:
counts_of_cluster: 26
Cluster 1:['platform', 'also', 'crowdfunding', 'mobile', 'equity', 'system', 'alternative', 'investor', 'online', 'fintechs']
Cluster 1:
counts_of_cluster: 67
Cluster 2:['indonesia', 'rural', 'msmes', 'financial', 'tourism', 'indonesian', 'china', 'emergence', 'fintech', 'internet']
Cluster 2:
counts_of_cluster: 6
Cluster 3:['islamic', 'malaysia', 'zakat', 'shariah', 'bcm', 'crowdfunding', 'variable', 'maqasid', 'big', 'governance']
Cluster 3:
counts_of_cluster: 11
Cluster 4:['india', 'startup', 'journal', 'inclusion', 'economy', 'system', 'also', 'support', 'international', 'romanian']
Cluster 4:
counts_of_cluster: 25
Cluster 5:['ghana', 'africa', 'exporter', 'scheme', 'lender', 'borrower', 'collateral', 'asymmetry', 'devt', 'problem']
Cluster 5:
counts_of_cluster: 6
Cluster 6:['china', 'chinese', 'system', 'technological', 'financialization', 'ecological', 'mediumsized', 'yuan', 'establish', 'environ
ment']
Cluster 6:
counts_of_cluster: 6
Cluster 7:['blockchain', 'financial', 'finance', 'trade', 'firms', 'transaction', 'eximchain', 'system', 'node', 'fintech']
Cluster 7:
counts_of_cluster: 10
Cluster 8:['think', 'design', 'fachhochschule', 'zürcher', 'stadelmann', 'fintechs', 'learn', 'team', 'alliance', 'digitalization']
Cluster 8:
counts_of_cluster: 8
Cluster 9:['sbvr', 'ontology', 'verb', 'brand', 'vocabulary', 'rule', 'concept', 'identity', 'mercury', 'xml']
Cluster 9:
counts_of_cluster: 3
```

From figure 33, we can see that most of the articles were assigned to each cluster 1. From figure to figure, the titles of each cluster is shown.

3.3.4 Hierarchical clustering

We now implement hierarchical clustering, and we will use the agglomerative approach and Dendrogram to visualize Hierarchical clusters.

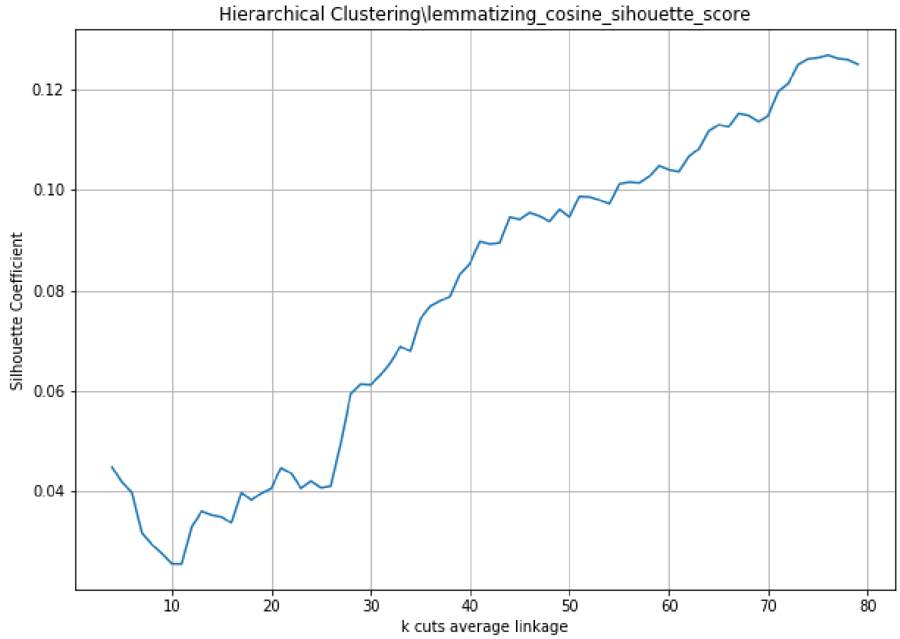
Agglomerative starts with each observation and merges the shortest two clusters with the linkage until there is only one large cluster left (Murtagh 1983).

The concept of linkage is important when dealing with hierarchical clustering. Linkage is used to calculate the dissimilarity in the groups. We can see that single-linkage clustering is used to compare the similarity of their most similar members, while the complete-linkage clustering is used to compare the similarity of their most dissimilar members. Average linkage is a compromise between single and complete linkage. If the dissimilarities between observations have a strong clustering tendency, then each linkage method produces relatively similar results, if not, then the method could produce very different results (Boehmke and Greenwell 2019). We will use average linkage and cosine distance. It is important to have an "elbow curve" for hierarchical clustering to choose the best "k." To differentiate it with K-means, we will call "k" for hierarchical clustering as "k cuts." We will introduce the concept of silhouette score, which is normally used to select "k cuts" for hierarchical clustering.

Silhouette score can measure the similarities of samples and their corresponding clusters. In other words, it is to measure the "cohesion". The range of the Silhouette score is between -1 and 1. If samples are strongly linked with their corresponding clusters, the score will be close to 1; otherwise, it is close to -1. If the score is close to 0, it means these samples are in the borderline of the clusters (Ogbuabor and Ugwoke 2018). We can see the result in figure 34.

We can conclude two characteristics from the figure 34. First, the silhouette coefficients gradually increase after "k cuts = 10". After a peak when "k cuts" equals 75 and it decreases. Since the number of our sample is 168, we will not take 75 as our best "k cuts" value. Secondly, although the overall figure shows an upward trend, all the values do not exceed 0.14 even at the peak. In other words, all the values are close to 0. Because we

Figure 34: Silhouette score for hierarchical clustering



already targeted our keywords when we scraped these articles from the internet, we know there are lots of similarities among these articles, and the result of these values is normal. To better compare these two methods, we will use "k=10" for hierarchical clustering when cluster our articles. And the labels is shown in figure 35.

Figure 35: Label class of hierarchical clustering

```
n_samples: 168
cluster 0, label ['china', 'university', 'point', 'start', 'production', 'file', 'crucial', 'school', 'abstract', 'yede']
counts_of_cluster: 21
cluster 1, label ['structure', 'medium', 'evidence', 'contribution', 'preference', 'file', 'abstract', 'award', 'winner', 'tun']
counts_of_cluster: 46
cluster 2, label ['factor', 'implementation', 'continuity', 'success', 'critical', 'structure', 'file', 'thesis', 'bin', 'submit']
counts_of_cluster: 11
cluster 3, label ['difficulty', 'asymmetric', 'global', 'analysis', 'university', 'disclosure', 'volume', 'file', 'school', 'cheng']
counts_of_cluster: 8
cluster 4, label ['ontology', 'brand', 'identity', 'abstract', 'university', 'file', 'paramount', 'today', 'faculty', 'advancement']
counts_of_cluster: 3
cluster 5, label ['journal', 'global', 'integrate', 'science', 'department', 'innovative', 'file', 'space', 'boost', 'print']
counts_of_cluster: 5
cluster 6, label ['strategy', 'global', 'meeting', 'community', 'advance', 'file', 'agenda', 'registration', 'banga', 'warren']
counts_of_cluster: 54
cluster 7, label ['speaker', 'science', 'music', 'scan', 'face', 'match', 'institute', 'example', 'agenda', 'file']
counts_of_cluster: 2
cluster 8, label ['equity', 'level', 'private', 'venture', 'play', 'association', 'file', 'minister', 'rock', 'magazine']
counts_of_cluster: 17
cluster 9, label ['expand', 'alternative', 'report', 'statistic', 'file', 'entrepreneur', 'horizon', 'federation', 'exchange', 'monthly']
counts_of_cluster: 1
```

Chapter 4

Results

4.1 Description of each cluster

We now analyse the result of hierarchical clustering.

We will analyze each cluster one by one, and then try to find the connection between the labels in each cluster with these articles. That is our way to prove how strong our labels can be used as "ground truth" for each cluster. Besides, we also hope that we can get the general idea of each cluster, and find the evidence that SMEs face the financial problems and the way that Fintech provides to address the SMEs financial problems.

4.2 Cluster 0

In cluster 0, there are 21 articles in this cluster, and we get the labels as follows: ['china', 'university', 'point', 'start', 'production', 'file', 'crucial', 'school', 'abstract', 'yede'].

In this cluster, we observe that the word 'china' has the strong link with these articles. Ma et al. (2010) describe the crucial role of SMEs in the Chinese economy. For example, it can promote employment and so on. At the same time, it is proposed that strengthening financial and technological innovation can promote the development of SMEs.

Gao and Gengzi (2019) and Ma et al.(2010) have the same views. These authors use various data to explain the impact of SMEs on the Chinese economy. In the meantime, these authors analyze the impact of the financial environment on SMEs from five different aspects, and at the same time, suggest that the existing imperfect financial environment is one of the important reasons for SMEs to get into financing difficulties.

Some authors also believe that the imperfect financial environment is one of the important reasons that cause the financing difficulties of SMEs. At the same time, according to the characteristics of Chinese corporate financing, some authors also propose countermeasures to address SME financing problems. Lizhen et al.(2014) propose the idea of "Blue Ocean Strategy" based on the characteristics of Chinese SMEs. With the collaboration of government and financial institutions, SMEs can solve the financing issues via "Blue Ocean Strategy."

Some authors have fully affirmed the level of innovation as an important driving force for the development of SMEs. For example, Huang and Zheng (2011) use Finland as a developed country and China as a developing country as an example to examine the critical driving forces of innovation and to analyze different innovation patterns based on theoretical and empirical studies. It describes innovation activities on the background of country-specific, especially identifying various factors determinant the innovativeness of SMEs.

Some articles, while fully affirming the economic dividends brought by innovations led by Fintech, have also mentioned China's special national context. Gruin and Julian (2019) think that digital credit scoring has made significant changes to the Chinese economy, for example, it has promoted financial liberalization, deepened financial inclusion, and developed Chinese economy. They critically examines the rise of China's Fintech and re-conceptualizes it as a tradition of the authoritarian rule of financialization embedded in the Chinese political system. This article is mainly related to the development of Fintech for strengthening the construction of China 's neo-statist authoritarian capitalism. The main idea of this article is focused on the political economy, economic sociology and Chinese studies.

In this cluster, in the context of China, the problems faced by Chinese SMEs were raised from all aspects and these authors fully affirm the contribution of Fintech to SMEs.

4.3 Cluster 1

In cluster 1, there are 46 articles in this cluster, and we get the labels as follows: ['structure', 'medium', 'evidence', 'contribution', 'preference', 'file', 'abstract', 'award', 'winner', 'tun'].

In this cluster, we know that this cluster provided a general view of problems faced by SMEs in different regions and the corresponding solutions for these financial problems. For example, in the southeast Asian region, Lin and Liu (2018) discussed that in the underdeveloped regions of Pan Asia, countries should formulate corresponding policies to take advantage of the rise of "disruptive technologies" while striving to improve the economies in the context of the global industrial division of labour. Lin and Liu (2018) point out that two disruptive technologies, such as blockchain and AI, have risk; they have their values in different fields.

Tambunan et al. (2019) mentioned that the number of SMEs in Indonesia had reached 99% of the total industries, and the SMEs have provided 92% of the total employment in this country. Although SMEs play an important role in the country's economic development, Tambunan et al. (2019) summarized three problems faced by SMEs in this country. First, SMEs have limited access to funds. Second, SMEs heavily depend on informal sources in this country. Third, banks are reluctant to make loans to SMEs due to the lack of collateral.

Nuryakin et al. (2019) and Prawirasasra et al. (2018) fully affirmed that Fintech had brought many benefits to Indonesia as a financing tool, but at the same time, they acknowledged that it is urgent to improve the laws and regulations of relevant regulatory agencies.

Rosavina et al. (2018) think that the peer to peer borrowing platform is alternative financing, which minimizes the obstacles when the borrowers have credit transactions with traditional banks and financial institutions. The peer to peer platform is also an important solution for SMEs who have difficulty obtaining financing. However, the number of loan transactions under Indonesia's P2P lending platform is low. Rosavina et al. (2018) found that the factors affecting SMEs' reluctance to resort to P2P platforms include loan process, interest rate, process cost, amount of loan, and Loan Application Flexibility. Rosavina et al. (2018) think that P2P companies should optimize their platforms in conjunction with government policies while providing assistance for the development of SMEs.

Hendriyani et al. (2019) think that technological development has promoted the development of financial technology literacy, and Fintech startups have entered the world of peer-to-peer lending to bridge the gap between traditional banking and those who cannot get funding from banks. This technology makes it easier, faster and cost-effective to access financing. In the age of digital finance in Indonesia, P2P loan companies have developed business agility strategies to attract their customers. By using this technology platform, they can simply gain a competitive advantage.

Moenjak et al. (2019) used data from 1.29 million SME loan contracts obtained from 15 Thai commercial banks and six professional financial institutions (SFI), and they found that only a few banks lend to SMEs on a relatively low loan scale. Although there are SFI platforms to assist SME loans, less than half of SMEs in Thailand obtain loans from these financial institutions. Based on the current background of Thailand, the author proposes a platform that uses the financial structure of the digital era to design SME loans.

Under the context of Rajkot region in India, Katrodia et al. (2018) also found that the financial institutions lack financial support for young entreprenuers. Different types of entrepreneurs face different levels of challenges in different business support services. Unlike other authors, Katrodia et al. (2018) also suggest that the education owned by young entrepreneurs influences their response to challenges.

Kwaning et al. (2015) urged the government in Ghana to formulate policies that favor SME financing. It also states that the SME Association must be formed to unite SMEs facing the same predicament and act as a guarantor when obtaining loans.

Abor et al. (2006) pointed out that SMEs can rely on common finance institutions besides conventional financial institutions.

Akorsu et al. (2012) mentioned that factors such as moral hazard are the main reasons to restrict credits. Akorsu et al. (2012) also proposed an alternative model to raise funds under the context of Ghana.

In this cluster, the authors propose different solutions in different regions according to different national conditions. if we link the contents to our labels, the words such as "structure" and "contribution" can have a good description for this cluster.

4.4 Cluster 2

In cluster 2, there are 11 articles and we found the labels as follows: ['factor', 'implementation', 'continuity', 'success', 'critical', 'structure', 'file', 'thesis', 'bin', 'submit'].

We found that these articles in this cluster also have regional characteristics. It mainly focuses on the issue of SME financing under the Islamic context. Ramlia et al. (2019) explore the concept Islamic finance in new technology based on Maqasid Syariah. Three issues are raised in issues in developing the new concept of Islamic financing in automobile industries. Firstly, Islamic automobile financing product awareness. Secondly there is conflict between conventional financing and Islamic Hire purchase financing product and new technology and maqasid syariah perspectives. Ramlia et al. (2019) think that as the world and technology are changing the new concept combining with Islamic contracts, technology must be based on Maqasid Syariah as the peer to peer lending, and crowdfunding needs a view from the Islamic perspective. The new financial technology and apps can be a big potential to develop new future Islamic automobile financing.

Other authors also mention crowdfunding can be used as an important means of financing in this region. For example, Saiti et al. (2018) provide an alternative model for financial markets. Combined with Fintech innovation, the problems can be addressed under the Islamic background.

Miskam et al. (2018) discussed that in the general context of the Islamic financial services sector, Fintech has in some ways changed the financial situation by promoting big data applications, and it has improved the efficiency of financial decisions. At the same time, Miskam et al. (2018) pointed out that in the context of Islamic financial services, by exploiting the value of information, it has been widely used in various fields such as financing, takaful, investment, and wealth management. Miskam et al. (2018) also think that big data can improve information sharing, operational efficiency, and customer experience.

Nasiren et al. (2017) state that Business Continuity Management (BCM) as an essential facilitates management strategic tool can help ensure the viability of any business in the event of a disruption and has been implemented in Malaysia for many years. But SMEs' knowledge of Business Continuity Management is still very low. SMEs also face difficulties in implementing BCM in the system. Nasiren et al. (2017) identified 16 critical success factors through research to determine how to implement BCM in SMEs in Malaysia.

In this cluster, it is mainly combined with the situation of Islam to explain the difficulties faced by SMEs and these authors provide different financing solutions for these industries.

4.5 Cluster 3

There are 8 articles in cluster 3 and the labels we found from this cluster are: ['difficulty', 'asymmetric', 'global', 'analysis', 'university', 'disclosure', 'volume', 'file', 'school', 'cheng'].

Anggreini et al. (2019) think that in the context of Indonesia's rural banks, Anggreini et al. (2019) used financial tools such as NPL, ROA, LDR and NLTA to compare whether the development of Fintech harms rural banks. Anggreini et al. (2019) found that in Indonesia, Fintech and rural banks have similar market segmentation in the fields of peer to peer lending for SMEs and the retail market. Although the emergence of Fintech makes people resort to Fintech companies, rural banks' liquidities are not harmed.

Huang et al. (2014) think that the root cause of financing difficulties for SMEs is the severe information asymmetry between financial institutions, which leads to adverse selection and moral hazard. Huang et al. (2014) think that the financing theory of SMEs, information asymmetry and credit rationing theory are used to analyze the financing difficulties of SMEs and crack financing difficulties.

Xu et al. (2009) also mentioned that information asymmetry is the main problem that restricts the development of SMEs (SMEs) financing. Xu et al. (2009) established a multi-level screening system to integrate financial and non-financial indicators.

By analyzing the articles in this cluster, we can conclude that words such as 'asymmetric' 'difficulty' can reflect the theme of this cluster. This cluster reflects that the difficulties encountered by SMEs in the financing process are mainly related to information asymmetry.

Other articles have also analyzed the problems that caused SMEs' financing difficulties. This cluster overlaps with the literature review. In the "Difficulties access to traditional financing" chapter of the literature review, we have already concluded the problems that caused the difficulties for SMEs, and we will not elaborate on them here.

4.6 Cluster 4

There are only three articles in cluster 4 and the labels that represent this cluster are as follows: label ['ontology', 'brand', 'identity', 'abstract', 'university', 'file', 'paramount', 'today', 'faculty', 'advancement'].

Elikan et al. (2018) stressed the importance of brand identity for startup and SMEs and to conceptualize brand identity in startups and SMEs, Elikan et al. (2018) have have proposed an ontology based on a systematic literature review, and the definition of relevant contextual elements. The ontology that Elikan et al. (2018) proposed is based on the OntoUML standards and it allows it to be mutable, expandable and reusable.

Ceci et al. (2016) and Butler et al. (2015) think that under the context of the financial industry in need of making its compliance assessment activities more effective, the financial institutions hope to use the methods provided by AI to address the problems caused by failing to comply with existing regulations, because the problems have caused the loss of billions of money. The problems can be addressed by regulatory compliance in the financial industry.

What the three articles have in common is how to solve the problems of SMEs. The first article is more different from the other two in the direction of establishing brand identity, while the latter two are from a more specific point of view. With the help of AI, companies can better comply with laws and regulations to expand the possibility of financing. From the observation of labels, we can summarize that words such as "brand," "identity," "ontology." can all reflect the characteristics of this cluster.

4.7 Cluster 5

There are five articles in cluster 5 and the labels are as follows: ['journal', 'global', 'integrate', 'science', 'department', 'innovative', 'file', 'space', 'boost', 'print']. Singhal et al. (2018) think that there is still much room for development for small and medium-sized enterprises in the cold chain industry in India. However, SMEs in the industry face the challenges of high initial investment despite government subsidies, country-specific regulations and taxes, and awareness of the management of perishable products.

Some authors have proposed the concept of supply chain financing. For example, Mou et al. (2018) think that supply chain finance has broken through the traditional credit model and rapidly developed into an innovative financial business discipline. Core companies have played a key role in improving supply chain financial credit. By analyzing the credit risk of core enterprises in supply chain finance, through the "fuzzy analytic hierarchy process" (FAHP), a supply chain financial credit risk assessment system was constructed to evaluate the credit risk of core enterprises quantitatively. This enables companies to take measures to control credit risk, thereby promoting the healthy development of supply chain finance.

Liu et al. (2014) think that supply chain financing is a useful and important way to reduce the financial burden on SMEs. Because it is pledged by inventory or part of the working capital of the financing enterprise and the supply chain financing breaks through the traditional mortgage model and reduces the urgent need of cash flows for loans to enterprises. The supply chain financing model increases the choice of financing methods for SMEs and greatly reduces the risks for business participants. With more commercial banks' participation and more emerging businesses, supply chain finance will be developed more extensively to satisfy SMEs and thus promote economic development.

These articles mentioned a new method to solve the financing problem of SMEs, which is to use supply chain financing.

4.8 Cluster 6

Cluster 6 accounts for the most articles, with a total of 54 articles. The labels are as follows: label ['strategy', 'global', 'meeting', 'community', 'advance', 'file', 'agenda', 'registration', 'banga', 'warren'].

Holotiuk et al. (2018) and Aagaard et al. (2015) acknowledge that SMEs are facing difficulties. Holotiuk et al. (2018) think, due to the shortcomings of banks and Fintech companies, an increase in alliances in the financial services industry can be observed. Aagaard et al. (2015) aim to identify and describe key internal business processes that contribute to growth through a longitudinal study of 11 SME case companies located in low-growth areas.

Costin et al. (2018) introduce the ever-changing situation of the banking industry, which involves the technological evolution and digitization process taking place in financial markets and society. Chia et al. (2019) mainly illustrate the examples of design thinking applied to the design of new products for banks and discuss opportunities and challenges. In other articles, the idea of Fintech has also been mentioned with different angles.Giudici et al. (2018) first explain the emerging technologies, such as blockchain or big data, contribute to the development of financial technology. At the same time, it is emphasized that while developing the corresponding technology, policymakers must improve the appropriate risk management methods. The development of artificial intelligence in the financial field will stimulate people to discuss risk management practices.

Nicoletti et al. (2017) emphasize that Fintech has made great progress in some areas, and also emphasizes that although Fintech has many benefits, it is also accompanied by factors such as overload data and extensive computation. Blockchains and cryptocurrencies are mentioned again as an important part of Fintech.

Kandpal et al. (2019) think that the cashless system is not only a necessity but also a need of today's order. Although there is strong policy support for new technologies, the usage of bank accounts and the uptake of formal financial services beyond savings accounts still face great difficulties. At the same time, amendments to the banking act and the establishment of trust between companies are also crucial to the development of the Indian economy.

Aruna et al. (2019) urged India to step up its efforts to transform India into a cashless economy through existing Fintech methods. Aruna et al. (2019) also point out that SMEs face obstacles in implementing technological innovation. This may hinder the development of Fintech innovation. At the same time, it is pointed out that with the rapid development of urbanization, the heavy use of the Internet and mobile phones, and the improvement of network security, India will provide development opportunities for the implementation of a cashless economy.

Singh et al. (2014) emphasize that micro-small and medium enterprises (MSMEs) have made a significant contribution to the Indian economy. The Indian market has developed rapidly. But at the same time, the Indian market faces pressure from international competition. And there is pressure on loans in the country. These are all important factors restricting the development of SMEs.

Boscor et al. (2015) think that for the development of export SMEs, quality, branding strateiges, innovation and government policy support will improve SME's competition in international markets.

Through the analysis of these articles in this cluster, it can be seen that it is difficult to analyze these articles from a unified perspective. But words such as "strategy," "global," can also see some characteristics of this cluster.

4.9 Cluster 7

There are two PPTs in cluster 7. The contents of the label are as follows. ['speaker', 'science', 'music', 'scan', 'face', 'match', 'institute', 'example', 'agenda', 'file']
Stadelmann et al. (2018) introduced the contents of data science and machine learning by ZHAW datalab. The contents are mainly an introduction to the institutions' contributions in the fields of robust deep learning, voice recognition, document analysis, and learning to learn & control. It also shows that the agency has made great progress in face matching, music scanning, speaker clustering, and learning to cluster. However, the specific contents of SMEs and Fintech are not involved. Because this cluster is an artifact, we will not consider these two PPTs.

4.10 Cluster 8

There are 17 articles in cluster 8 and labels are as follows: ['equity', 'level', 'private', 'venture', 'play', 'association', 'file', 'minister', 'rock', 'magazine']
Under the economies of transition countries in Central and Eastern Europe, Hyz et al. (2008) describe that the importance of the SME sector is related to economic democratization, social stability brought about by the development of the middle class, and the creation of a large number of new jobs. Hyz et al. (2008) also point out the fact that leasing has had a considerable impact on the development of the SME sector in the Polish economy. The analysis performed is based on professional literature, information from the leasing department, and opinions from leading market experts. Hyz et al. (2008) describe the structure of the lease, its potential development direction, and the main obstacles to the development of the essential types of lease transactions.
Carmona et al. (2018) point out that more and more Fintech services provided by newly established startups, traditional financial institutions and large technology companies can bring new competitive challenges to the competitive environment. Certain factors may lead to anti-competitive behaviours, and the behaviours are network effects that result

from using online platforms, accessing customer data, standardization, interoperability, and using algorithms. Combined with item-by-item service analysis, the study provides descriptive analysis and normative tools to predict and manage the occurrence of anti-competitive behaviors.

Vidovi et al. (2018) point out that emerging and small companies are often behind innovations that are vital to economic growth, especially in software, nanotechnology, biotechnology, and clean technology. Blockchain technology and digital financial technology services are likely to completely change the financial market, especially the way to solve information asymmetry. SMEs are often the most vulnerable to the shortage of collateral. Other emerging technologies, such as industrial robotics, biotechnology, 3D printing, nanotechnology, also heavily involve the distribution and production of SMEs. Vidovi et al. (2018) also point out that knowledge spillovers, access to networks and opportunities to partner with larger enterprises have influenced SME's innovation. Potential synergies and trade-off taking place across diverse areas of interaction in the policy mix, including distortionary effects that may be introduced by some policy actions, should be paid more attention.

At the same time, Vidovi et al. (2018) proposed that the connection between local, national and global levels is also a challenge for SMEs, but the establishment of this entrepreneurial eco-system can facilitate innovative entrepreneurial spirit in different economic environments.

Vasilescu et al. (2010) propose the concept of factoring, which is a complete financial package that combines working capital financing, credit risk protection, accounts receivable bookkeeping and collection services. For the financial needs of SMEs and start-ups, Factoring can replace traditional bank credit because the advantage of factoring is that the SMEs can cash its debt. Based on the fact that factoring does not require collateral, it can enable SMEs to improve cash flow without increasing debt levels.

Based on the Serbian example, Eric et al. (2011) believe that although innovative SMEs are the backbone of innovation in the development of economies, these SMEs face difficulties in financing. Eric et al. (2011) think that undeveloped financial markets, unfavourable terms of debt financing and insufficient volume and scope of government support programs cause the financing problems for SMEs.

4.11 Cluster 9

There is one article in cluster 9 and the labels are as follows: ['expand', 'alternative', 'report', 'statistic', 'file', 'entrepreneur', 'horizon', 'federation', 'exchange', 'monthly'].
Ivanets et al. (2018) reviewed more than 200 articles of the theme of Fintech, and they mainly summarize the following points of Fintech by statistics. Firstly, large investment banks have conducted transactions or "exits" with a larger amount of money. Second, from the perspective of traditional vertical and hubs, Fintech's innovation is more diverse. Third, in addition to big banks and large insurance companies, mid-tier banks, insurance and wealth management also actively participate in corporate investments. Fourth, the new policies help the development of Fintech investment. Because this article is more similar to a survey, and that is why it is different from other clusters.

4.12 Overview of results

The thesis mainly discussed the use of K-means and hierarchical clustering for document clustering. At the same time, by reviewing 168 articles in ten clusters, we find that hierarchical clustering is more effective than K-means clustering for our dataset.
Because our dataset is based on two sets of keywords which ara "SME financing difficulties" and "Fintech SME solutions," these articles in the clusters have great similarities. We can also know the strong similarities when we see the silhouette score. The silhouette score for "k cut" is around 0.
By reviewing these articles, we can also conclude that some clusters show strong dissimilarities with other clusters. Cluster 4, cluster 7 and cluster 9 have strong dissimilarities with other clusters. In cluster 7, the two PPTs are mainly about the applications of data

science in different fields and the progress led by ZHAW datalab. However, these two PPTs did not cover much information about SMEs and Fintech.

The reason why Cluster 4 is different from other clusters is mainly that these three articles in cluster 4 are all about building models to solve specific problems for SMEs, such as the brand identity for SMEs and regulatory compliance for SMEs. These authors also provided models associated with different problems. The articles in other clusters tend to understand the difficulties faced by SME from a macro perspective and how to solve SMEs' problems by including Fintech and other means. The article in cluster 9 uses statistics to explain the expanding influence of Fintech in the field of financial services.

In terms of cluster similarity, according to SMEs' problems of financing difficulties. Cluster 0, cluster 1, cluster 2, cluster 3 and cluster 8 all summarize the reasons. In cluster 3, one of the prominent reasons that SMEs face financing difficulties is mainly due to the information asymmetry between banks and SMEs.

For cluster 0, cluster 1, cluster 2 and cluster 8, we can also see strong similarities with each other. These clusters try to discuss the SME financing difficulties under the context of underdeveloped areas. cluster 0 is based on China, and cluster 1 is mainly based on the underdeveloped countries in Southeast Asia, South Asia, and Africa. Cluster 2 is based on Islamic countries, and Cluster8 is based on some Eastern European countries. The geographical characteristics of these countries are different, so the focus of solving the financing problems of SMEs is different. For example, in cluster 1, those authors mentioned that it is important to establish an SME alliance to combat financing problems. Cluster 2 emphasizes that when dealing with SME financing problems, policymakers should consider Islamic background.

Besides the solutions presented by cluster 0, cluster 1, cluster 2 and cluster 8 based on the different regions, cluster 5 and cluster 6 also give different perspectives. The authors in cluster 5 think that the method of supply chain financing is an alternative solution, and the articles in cluster 6 are inclined to address the SME financing problems with the development of disruptive technologies.

By analyzing the correlation of these clusters, we can conclude that SMEs play an important role in both developing and developed countries. However, problems such as information asymmetry between banks and SMEs, SMEs-unfriendly policies hinder the SME financing from the official source. Resolving the issue of policies is an important prerequisite. Besides, the development of disruptive technologies such as AI and blockchain can help solve SME financing problems. Other than that, it is also helpful to take advantage of the SME alliance and supply chain financing to solve the problem for SMEs.

Chapter 5

Conclusion

An important task in research consists in analyzing the literature. However, since the number of publications tend to increase with time, this task may get difficult. Various approaches exist to reduce the number of references for a given topic. The first consists in narrowing the research by adding keywords or by combining a basic search with filters on citations or citing articles. Unfortunately, these approaches usually exclude papers that could be of interest.

Unfortunately, adding keywords finally results in narrowing the search in a way that is not always appropriate. Some papers may be of interest, even if they do not fit the first idea of the researcher.

Instead of this a priori approach, in this thesis, we use an alternative technique which consists in analyzing automatically the papers and extract information from their content by the mean of text mining. In this way, we could make groups of papers that are dealing with the same topic and have a better idea of the trend in research on that topic. As opposed to the keyword approach, we deduce the topics from the papers and they are not restricted by the researcher. In this way, new trends may be considered that would otherwise be missed. Technically, by only using simple keywords such as “SME financing difficulties” and “Fintech SME solution”, almost 200 papers were obtained. A first rough analysis was used to identify some papers that were difficult to analyze altogether with the others, mainly because they were written in a different language. A second step in the analysis was to apply text mining in order to group those papers in few clusters. In this case, 10 clusters were selected.

A deeper analysis allows us to identify some interesting trends like the relation between SME financing through fintech and cryptomoney, or the special case of financing SMEs in the special context of china. Obviously, identifying these groups of papers that prove a special interest of the scientific community would not be possible if specific keywords were used.

From the financing SMEs point of view, the main achievement of this thesis is the identification of topics that could have otherwise been missed. The next step for a researcher is to consider only a group of those papers based upon their content and focus on the so found rather small number of relevant papers.

From a more general point of view, we prove that the principle of meta-analysis may be considered in a simple way by the use of computer aided methods like text mining. As opposed to more classical approaches for making a literature review, this approach is highly scalable and allows the researcher to implicitly consider a large number of articles.

Bibliography

Aagaard, A. and Senderovitz, M. (2015). How can smes support growth in a low-growth region?: A longitudinal study of sme growth facilitation. *Danish Journal of Management and Business*, 79(2):21–34.

Abdulrahman, R., Neagu, D., Holton, D. R. W., Ridley, M., and Lan, Y. (2013). *Data Extraction from Online Social Networks Using Application Programming Interface in a Multi Agent System Approach*, pages 88–118. Springer Berlin Heidelberg, Berlin, Heidelberg. doi: 10.1007/978-3-642-41776-4_4.

Abor, J. and Biekpe, N. (2006). Small business financing initiatives in ghana. *Problems and Perspectives in Management*, 4(3):69–77.

Ahuja, M., Singh, J., and nica, V. (2014). Web crawler: Extracting the web data. *International Journal of Computer Trends and Technology*, 13:132–137. doi: 10.14445/22312803/IJCTT-V13P128.

Akaichi, J., Dhouioui, Z., and Pérez, M. J. L.-H. (2013). Text mining facebook status updates for sentiment classification. In *2013 17th International conference on system theory, control and computing (ICSTCC)*, pages 640–645.

Akorsu, P. K. and Agyapong, D. (2012). Alternative model for financing smes in ghana. *International Journal of Arts and Commerce*, 1(5):136–148.

Amudha, S. and Phil, M. (2017). Web crawler for mining web data. *International*

Research Journal of Engineering and Technology, 4(02). Retrieved from https://www.irjet.net/archives/V4/i2/IRJET-V4I225.pdf.

Anggreini, S. I. et al. (2019). The disruption of fintech on rural bank: An empirical study on rural banks in indonesia. Retrieved from https://www.researchgate.net/publication/336891020_The_Disruption_of_Fintech_On_Rural_Bank_An_Empirical_Study_On_Rural_Banks_in_Indonesia.

Arner, D. W., Barberis, J., and Buckey, R. P. (2016). Fintech, regtech, and the reconceptualization of financial regulation. *Nw. J. Int'l L. & Bus.*, 37:371.

Aruna, P. and Acharya, S. (2019). Fintech: Ushering in a digital revolution. *Navajyoti, International Journal of Multi-Disciplinary Research*, 3.

Bengfort, B., Bilbro, R., and Ojeda, T. (2018a). Building a custom corpus. In *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*, pages 47–71. O'Reilly Media.

Bengfort, B., Bilbro, R., and Ojeda, T. (2018b). Text vectorization and transformation pipelines. In *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*, pages 88–118. O'Reilly Media.

Berger, A. and Udell, G. (1995). Universal banking and the future of small business lending. *Wharton School Center for Financial Institutions, University of Pennsylvania, Center for Financial Institutions Working Papers*, (No. FIN-95-009).

Berger, A. N. and Frame, W. S. (2007). Small business credit scoring and credit availability*. *Journal of Small Business Management*, 45(1):5–22. doi: 10.1111/j.1540-627X.2007.00195.x.

Berger, A. N. and Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal*, 112(477):F32–F53. doi: 10.1111/1468-0297.00682.

Berger, A. N. and Udell, G. F. (2006). A more complete conceptual framework for sme finance. *Journal of Banking Finance*, 30(11):2945–2966. doi: https://doi.org/10.1016/j.jbankfin.2006.05.008.

Bholowalia, P. and Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).

Boehmke, B. and Greenwell, B. M. (2019). *21 Hierarchical Clustering*. Retrieved from https://bradleyboehmke.github.io/HOML/hierarchical.html#hierarchical-clustering-algorithms.

Bolasco, S., Canzonetti, A., Capo, F., Della Ratta, F., and Singh, B. (2006). *Understanding Text Mining: A Pragmatic Approach*, volume 185, pages 31–50. doi: 10.1007/3-540-32394-5_4.

Boscor, D. (2015). Export strategies adopted by romanian sme. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, 8(1):227–299.

Boskov, T. (2018). Level of deepening financial infrastructure, fintech companies and financial inclusion: Theory and evidence. *IJIBM International Journal of Information, Business and Management*, 10(4):23–31.

Butler, T., Abi-Lahoud, E., O'Brien, L., Ceci, M., Emerson, B., and Naz, T. (2015). Knowledge elicitation for regulatory compliance in the financial industry. *GRCTC WORKING PAPER June 2015*. Retrieved from https://e9a5d5c6.stackpathcdn.com/wp-content/uploads/2015/12/GR3C-Knowledge-Elicitation-for-Regulatory-Compliance.pdf.

CARMONA, A. F., LOMBARDO, A. G.-Q., Rivera, R., Pastor, C., GARCÍA, J. V., MUÑOZ, D. R., and MARTÍN, L. C. (2018). Competition issues in the area of financial technology (fintech). *Policy Department for Economic, Scientific and Quality of Life Policies of the European Parliament*, PE 619.027 - July 2018. Retrieved from http://www.quanticae.com/wp-content/uploads/2018/07/IPOL_STU2018619027_EN.pdf.

Ceci, M., Al Khalil, F., and O'Brien, L. (2016). Making sense of regulations with sbvr. In *RuleML (Supplement)*. Retrieved from http://ceur-ws.org/Vol-1620/paper7.pdf.

Chia, A. J. H. and Lee, J.-J. (2019). Banking outside-in: How design thinking is changing the banking industry? *CfP: IASDR 2019 Conference – Design Revolutions*. Retrieved from https://iasdr2019.org/uploads/files/Proceedings/ch-f-1263-Chi-A.pdf.

Cooley, R., Mobasher, B., and Srivastava, J. (1997). Web mining: information and pattern discovery on the world wide web. In *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*, pages 558–567. doi: 10.1109/TAI.1997.632303.

Costa, M. G. d. and Zhiguo, G. (2005a). Web structure mining: an introduction. In *2005 IEEE International Conference on Information Acquisition*, page 6. doi: 10.1109/ICIA.2005.1635156.

Costa, M. G. d. and Zhiguo, G. (2005b). Web structure mining: an introduction. In *2005 IEEE International Conference on Information Acquisition*, page 6. doi: 10.1109/ICIA.2005.1635156.

Cuesta, A., Barrero, D., and R-Moreno, M. (2014). A framework for massive twitter data extraction and analysis. *Malaysian Journal of Computer Science*, 27:50–67.

Dai Yueming, W. M., Ming, Z., and Yan, W. (2018). Optimizing initial cluster centroids by svd in k-means algorithm for chinese text clustering. *Journal of System Simulation*, 30(10):3835.

Dang, S. (2014). Text mining : Techniques and its application. *International Journal of Enginerring Technology Innnovation*, 1:22–25.

El Zahra, F., Zafar, W. U. I., Zafar, F., and Ayesha, S. (2010). Impact and usage of internet in educational community in pakistan. *European Journal of Scientific Research*, 47:256–264.

Elikan, D. and Pigneur, Y. (2018). Brand identity ontology. In *VMBO*, pages 133–141.

Eric, D., Beraha, I., and Djuricin, S. (2011). Financing innovative small and mediumsized enterprises in times of crisis. *Revista Romana de Economie*, 33(2).

Fenwick, M. and Vermeulen, E. P. (2017). How to respond to artificial intelligence in fintech. *Japan SPOTLIGHT*, pages 16–20.

Fu, F., Chen, X., Liu, L., and Wang, L. (2007). Social dilemmas in an online social network: The structure and evolution of cooperation. *Physics Letters A*, 371:58–64. doi: `10.1016/j.physleta.2007.05.116`.

Gao, G. (2019). Research on the impact of financial ecological environment on sme financing. *Ekoloji Dergisi*, (107):1633–1640.

Getoor, L. (2003). Link mining: A new data mining challenge. *SIGKDD Explorations*, 5:84–89. doi: `10.1145/959242.959253`.

Giudici, P. (2018). Fintech risk management: A research challenge for artificial intelligence in finance. *Frontiers in Artificial Intelligence*, 1(1):1–6.

Grey Literature International Steering Committee (2006). Guidelines for the production of scientific and technical reports: how to write and distribute grey literature. Retrieved from `http://eprints.rclis.org/7469/1/nancy.pdf`.

Gruin, J. (2019). Financializing authoritarian capitalism: Chinese fintech and the institutional foundations of algorithmic governance. *Finance and Society*, 5(2):84–104.

Haddaway, N. (2015). The use of web-scraping software in searching for grey literature. *Grey Journal*, 11:186–190.

Hendriyani, C. and Raharja, S. J. (2019). Business agility strategy: Peer-to-peer lending of fintech startup in the era of digital finance in indonesia. *Review of Integrative Business and Economics Research*, 8:239–246.

Holotiuk, F., Klus, M. F., Lohwasser, T. S., and Moormann, J. (2018). Motives to form alliances for digital innovation: The case of banks and fintechs. In *Bled eConference*, page 22.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, volume 4, pages 9–56.

Huang, C., When, Y., and Liu, Z. (2014). Analysis on financing difficulties for smes due to asymmetric information. *Global Disclosure of Economics and Business*, 3(2):28–31.

Huang, Z. (2011). Sme strategic management and innovation—a comparative study between finland and china. *Proceedings of the Sustainable Innovation, Entrepreneurship and Growth*, pages 15–18.

Hyz, A. and Gikas, G. (2008). The development of sme sector in poland and new instruments of financing. In *2nd International Conference on Business and Management Education*, pages 17–20.

Ibrahim, N., Hassan, A., and Nihad, M. (2018). Big data analysis of web data extraction. *International Journal of Engineering and Technology*, 7:168–172. doi: 10.14419/ijet.v7i4.37.24095.

Inamdar, S. and Shinde, G. (2008). An agent based intelligent search engine system for web mining. *Research, Reflections and Innovations in Integrating ICT in education*.

Industry Canada (2013). Small business branch financing statistics special edition: Key small business statistics november 2013. Retrieved from https://www.ic.gc.ca/eic/site/061.nsf/vwapj/FinancingKSBS-FinancementPSRPE_2013_eng.pdf/

\protect\T1\textdollarFILE/FinancingKSBS-FinancementPSRPE_2013_eng.pdf.

Innovation, Science and Economic Development Canada (2019). Key small business statistics january 2019. Retrieved from https://www.ic.gc.ca/eic/site/061.nsf/vwapj/KSBS-PSRPE_Jan_2019_eng.pdf/\protect\T1\textdollarfile/KSBS-PSRPE_Jan_2019_eng.pdf.

Ivanets, (2018). Fintech's growing influence on financial services. Retrieved from https://ir.kneu.edu.ua/bitstream/handle/2010/26043/KF_2018_12.pdf?sequence=1.

Kallberg, J. G. and Udell, G. F. (2003). The value of private sector business credit information sharing: The us case. *Journal of Banking Finance*, 27(3):449–469. doi: https://doi.org/10.1016/S0378-4266(02)00387-4.

Kanakaraj, M. and Guddeti, R. M. R. (2015). Nlp based sentiment analysis on twitter data using ensemble classifiers. In *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*, pages 1–5.

Kandpal, V. and Mehrotra, R. (2019). Financial inclusion: The role of fintech and digital financial services in india. *Indian Journal of Economics & Business*, 19(1):85–93.

Katrodia, A. and Sibanda, R. (2018). Challenges experienced by young entrepreneurs entering the small and medium enterprise sector (sme) in rajkot region, gujarat-india. *Academy of Entrepreneurship Journal*, 24(4):1–14.

Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. 46(5):668–677. doi: https://doi.org/10.1145/324133.324140.

Kobayashi, M. and Aono, M. (2008). Vector space models for search and cluster mining. In *Survey of Text Mining II*, pages 109–127. Springer.

Kosala, R. and Blockeel, H. (2000a). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1):1–15.

Kosala, R. and Blockeel, H. (2000b). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1):1–15. doi: https://doi.org/10.1145/360402.360406.

Kumar, A. and Singh, R. K. (2016). Web mining overview, techniques, tools and applications: A survey. *International Research Journal of Engineering and Technology (IRJET)*, 3(12):1543–1547.

Kwaning, C. O., Nyantakyi, K., and Kyereh, B. (2015). The challenges behind smes' access to debts financing in the ghanaian financial market. *International Journal of Small Business and Entrepreneurship Research*, 3(2):16–30.

Lee, S. (2017). Evaluation of mobile application in user's perspective: Case of p2p lending apps in fintech industry. *TIIS*, 11(2):1105–1117.

Lin, C.-F. and Liu, H.-W. (2018). Disruptive technologies and sustainable development: Implications for southeast asia. *International Centre for Trade and Sustainable Development (ICTSD)*. Retrieved from https://www.ictsd.org/sites/default/files/research/disruptive_technologies_and_southeast_asia_-_liu_and_lin.pdf.

Liu, W. (2014). Boosting smes with better and innovative financing services. *Journal of System and Management Sciences*, 4(2):15–23.

Lizhen, M. (2014). Sme financing of the post-crisis era in china based on blue ocean strategy. *Guangxi University of Finance and Economics, Guangxi, P.R.China*. Retrieved from http://www.seidatacollection.com/upload/product/201309/2013gjmy325f3.pdf.

Lomborg, S. and Bechmann, A. (2014). Using apis for data collection on social media. *The Information Society*, 30. doi: 10.1080/01972243.2014.915276.

Ma, J., Wang, S., and Gui, J. (2010). A study on the influences of financing on technological innovation in small and medium-sized enterprises. *International Journal of Business and Management*, 5(2):209.

Magnuson, W. (2018). Regulating fintech. *Vand. L. Rev.*, 71:1167.

Maletic, J. I. and Valluri, N. (1999). Automatic software clustering via latent semantic analysis. In *14th IEEE International Conference on Automated Software Engineering*, pages 251–254.

Matthew, A. (2017). Crowd-sourced equity funding: The regulatory challenges of innovative fintech and fundraising. *U. Queensland LJ*, 36:41.

Miller, M. J. (2003). Credit reporting systems around the globe: The state of the art in public credit registries and private credit reporting firms. *Credit Reporting Systems and the International Economy*, pages 25–80.

Mills, K. G. and McCarthy, B. (2016). The state of small business lending: Innovation and technology and the implications for regulation. Harvard Business School Working Paper 17-042. doi: https://doi.org/10.2139/ssrn.2877201.

Miskam, S. and Eksan, S. H. R. (2018). Big data and fintech in islamic finance: Prospects and challenges. *4th Muzakarah Fiqh International Fiqh Conference (MFIFC 2018)*, pages 236–244. Retrieved from http://conference.kuis.edu.my/mfifc/images/e-proceeding/2018/236-244.pdf.

Moenjak, T., Praneeprachachon, V., Bumpenboon, T., Bumrungruan, P., Monchaitrakul, C., et al. (2019). Gaining from digital disruption: the thai financial landscape in the digital era. Technical report, Puey Ungphakorn Institute for Economic Research.

Mou, W., Wong, W.-K., and McAleer, M. (2018). Financial credit risk evaluation based on core enterprise supply chains. *Sustainability*, 10(10):3699.

Mughal, M. J. H. (2018). *Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview*, volume 9. (IJACSA) International Journal of Advanced Computer Science and Applications.

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4):354–359.

Nakamura, L. (1994). Small borrowers and the survival of the small bank: is mouse bank mighty or mickey? *Business Review*, pages 3–15.

NASIREN, M. D. B. (2017). A structured critical success factors model for business continuity management implementation in malaysia smes. *UNIVERSITI TEKNOLOGI MALAYSIA*. Retrieved from `http://eprints.utm.my/id/eprint/78852/1/MuhamadDanialAtifMFGHT2017.pdf`.

Nicoletti, B., Nicoletti, and Weis (2017). *Future of FinTech*, volume Fintech Innovation. Springer.

NIŢESCU, D. C. and DUNA, F. A. (2018). Managing digitalization in banking: Challenges and implications. *Management Perspectives in the Digital Era*. Retrieved from `http://conferinta.management.ase.ro/archives/2018/pdf/2_17.pdf`.

Nuryakin, C., Aisha, L., Massie, N. W. G., et al. (2019). Financial technology in indonesia: A fragmented instrument for financial inclusion? Technical report, LPEM, Faculty of Economics and Business, University of Indonesia.

Ogbuabor, G. and Ugwoke, F. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *International Journal of Computer Science & Information Technology*, 10(2):27–37.

Omarova, S. T. (2019). New tech v. new deal: Fintech as a systemic phenomenon. *Yale J. on Reg.*, 36:735.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Report, Stanford InfoLab.

Pons-Porrata, A., Berlanga-Llavori, R., and Ruiz-Shulcloper, J. (2007). Topic discovery based on text mining techniques. *Information Processing Management*, 43(3):752–768. doi: https://doi.org/10.1016/j.ipm.2006.06.001.

Prawirasasra, K. P. (2018). Financial technology in indonesia: disruptive or collaborative. *Reports on Economics and Finance*, 4(2):84.

Ramlia, M. A. R., Basarya, N. F., and Anuar, M. K. (2019). The concept of future islamic automobile financing based on maqasid shariah. *RC2019- Research Colloquium 2019*, 1(2019):22–25.

Rosavina, M. and Rahadi, R. A. (2018). Peer-to-peer (p2p) lending platform adoption for small medium enterprises (smes): a preliminary study. *International Journal of Accounting*, 3(10):1–14.

Sachin, M. and Divya, G. (2018). An introduction to part-of-speech tagging and the hidden markov model. Retrieved from https://www.freecodecamp.org/news/an-introduction-to-part-of-speech-tagging-and-the-hidden-markov-model-953d45338f24/.

Saiti, B., Musito, M. H., and Yucel, E. (2018). Islamic crowdfunding: fundamentals, developments and challenges. 62(3). Retrieved from https://openaccess.izu.edu.tr/xmlui/bitstream/handle/20.500.12436/861/Islamic_Crowdfunding_Fundamentals_Developments_and_Challenges.pdf?sequence=1&isAllowed=y.

Saksonova, S. and Kuzmina-Merlino, I. (2017). Fintech as financial innovation–the possibilities and problems of implementation. *European Research Studies*, 20(3A):961.

Sasirekha, K. and Baby, P. (2013). Agglomerative hierarchical clustering algorithm-a. *International Journal of Scientific and Research Publications*, 83:83.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., and Lin, C.-T. (2017a). A review of clustering techniques and developments. *Neurocomputing*, 267:664–681.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., and Lin, C.-T. (2017b). A review of clustering techniques and developments. *Neurocomputing*, 267:664–681. doi: https://doi.org/10.1016/j.neucom.2017.06.053.

scikit-yb developers, T. (2019). *Yellowbrick*. Retrieved from https://www.scikit-yb.org/en/latest/api/cluster/elbow.html.

Shen, Z., Krimpen, H., and Spruit, M. (2019). A lightweight api-based approach for building flexible clinical nlp systems. *Journal of Healthcare Engineering*, 2019:1–11. doi: 10.1155/2019/3435609.

Shi, J. and Yang, R. (2009). The application of lending technology in sme financing. *Central University of Finance and Economics*.

Singh, J. and Gupta, V. (2017). A systematic review of text stemming techniques. *Artificial Intelligence Review*, 48(2):157–217.

Singh, J. and Singh, D. (2014). Problems related to the financing of small firms in india. *International Journal of Innovative Research and Development*, 3(1):317–321.

Singhal, V. and Koul, S. (2018). Identification of benchmarks for adoption by smes in the indian cold chain. *Fifteenth AIMS International Conference on Management*, pages 209–214. Retrieved from http://www.aims-international.org/aims15/15ACD/PDF/A209-Final.pdf.

Srivastava, J., Cooley, R., Deshpande, M., and Tan, P.-N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, 1(2):12–23.

Srivastava, J., Desikan, P., and Kumar, V. (2002). Web mining: Accomplishments and future directions. In *National Science Foundation Workshop on Next Generation Data Mining (NGDM'02)*, pages 51–56.

Stadelmann, T. (2018). Data science@ zhaw. Retrieved from `https://stdm.github.io/downloads/talks/2018-07-10_Data-Science-at-ZHAW.pdf`.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5):1891–1921. doi: `10.1111/0022-1082.00483`.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410. Retrieved from `http://www.jstor.org/stable/1802787`.

Sukhija, K. (2015). Web content mining equipped natural language processing for handling web data. *International Journal of Computer Applications Technology and Research*, 4(3):209–213.

Tambunan, T. (2019). Indonesian small businesses and their access to financing. *International Journal of Business Management and Economic Review*, 02:88–105.

Temelkov, Z. and Gogova Samonikov, M. (2018). The need for fintech companies as nonbank financing alternatives for sme in developing economies. *International Journal of Information, Business and Management*, 10(3):25–33.

Thelwall, M. (2001). A web crawler design for data mining. *Journal of Information Science*, 27:319–325. doi: `10.1177/016555150102700503`.

Ting, I. H. (2008). Web mining techniques for on-line social networks analysis. In *2008 International Conference on Service Systems and Service Management*, pages 1–5. doi: `10.1109/ICSSSM.2008.4598506`.

Vargiu, E. and Urru, M. (2013). Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artificial Intelligence Research*, 2. doi: 10.5430/air.v2n1p44.

Vasilescu, L. G. (2010). Factoring–financing alternative for smes. In *5th Annual MIDES International Conference, Kavala*, pages 13–23.

VIDOVI, M. T. (2018). Creation of support ecosystem for innovative fast growing companies. *SYSTEMICS, CYBERNETICS AND INFORMATICS*, 16(6). Retrieved from http://www.iiisci.org/journal/CV\protect\T1\textdollar/sci/pdfs/MA147HZ18.pdf.

Williamson, O. E. (1967). Hierarchical control and optimum firm size. *Journal of Political Economy*, 75(2):123–138. Retrieved from http://www.jstor.org/stable/1829822.

Wonglimpiyarat, J. (2017). Fintech banking industry: a systemic approach. *foresight*.

Wright, C. (2016). Fintech: Blockchain landmarks point to transformation of trade finance. Retrieved from https://www.euromoney.com/article/b12kq752dh74gj/fintech-blockchain-landmarks-point-to-transformation-of-trade-finance.

XU, K. and YANG, Y. (2009). Enterprise financing modeling based on information asymmetry. *2009 International SME Development Forum*.

Xu, Y. (2017). The literature review of text data mining. *Science Discovery*, 5:438. doi: 10.11648/j.sd.20170506.18.

Yoshida, G. (2019). Competition where banks and fintech companies are stronger together. *The RMA Journal*, pages 48–53.