

IMPLEMENTATION OF DATA MINING IN PT. X FOR BUSINESS ANALYSIS

A THESIS

**Submitted in Partial Fulfillment of the Requirement for the Degree of
Bachelor of Engineering in Industrial Engineering**



CLAUDIUS ANDIKA DANESWARA

16 14 08830

INTERNATIONAL INDUSTRIAL ENGINEERING PROGRAM

DEPARTMENT OF INDUSTRIAL ENGINEERING

FACULTY OF INDUSTRIAL TECHNOLOGY

UNIVERSITAS ATMA JAYA YOGYAKARTA

YOGYAKARTA

2020

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

IMPLEMENTATION OF DATA MINING IN PT. X FOR BUSINESS ANALYSIS

yang disusun oleh

CLAUDIUS ANDIKA DANESWARA

161408830

dinyatakan telah memenuhi syarat pada tanggal 04 Agustus 2020

		Keterangan
Dosen Pembimbing 1	: Kristanto Agung Nugroho, S.T., M.Sc.	Telah menyetujui
Dosen Pembimbing 2	: Kristanto Agung Nugroho, S.T., M.Sc.	Telah menyetujui
Tim Penguji		
Penguji 1	: Kristanto Agung Nugroho, S.T., M.Sc.	Telah menyetujui
Penguji 2	: Dr. Yosephine Suharyanti, S.T., M.T.	Telah menyetujui
Penguji 3	: Ririn Diar Astanti, D.Eng.	Telah menyetujui

Yogyakarta, 04 Agustus 2020

Universitas Atma Jaya Yogyakarta

Fakultas Teknologi Industri

Dekan

ttd

Dr. A. Teguh Siswanto, M.Sc

DECLARATION OF ORIGINALITY OF RESEARCH

I certify that the research entitled "Implementation of Data Mining in PT. X For Business Analysis" in this thesis has not already been submitted for any other degree.

I certify that to the best of my knowledge and belief, this thesis which I wrote does not contain the works of parts of the works of other people, except those cited in the quotations and bibliography, as a scientific paper should.

In addition, I certify that I understand and abide the rule stated by the Ministry of Education and Culture The Republic of Indonesia, subject to the provisions of Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Signature :



Student name : Claudius Andika Daneswara

Student ID : 161409038

Date : 17th July 2020

ACKNOWLEDGMENTS

The author expresses his sincere gratitude to his supervisor, Kristanto Agung Nugroho, S.T., M.Sc., for providing much meaningful advice, comments, and suggestions during his study and the preparation of this thesis.

His appreciation is also given to Ririn Diar Astanti, D.Eng, who serves as Head of Industrial Engineering Program.

Last but not least, sincere gratitude addressed to his beloved friends likewise family within International Industrial Engineering Study Program.

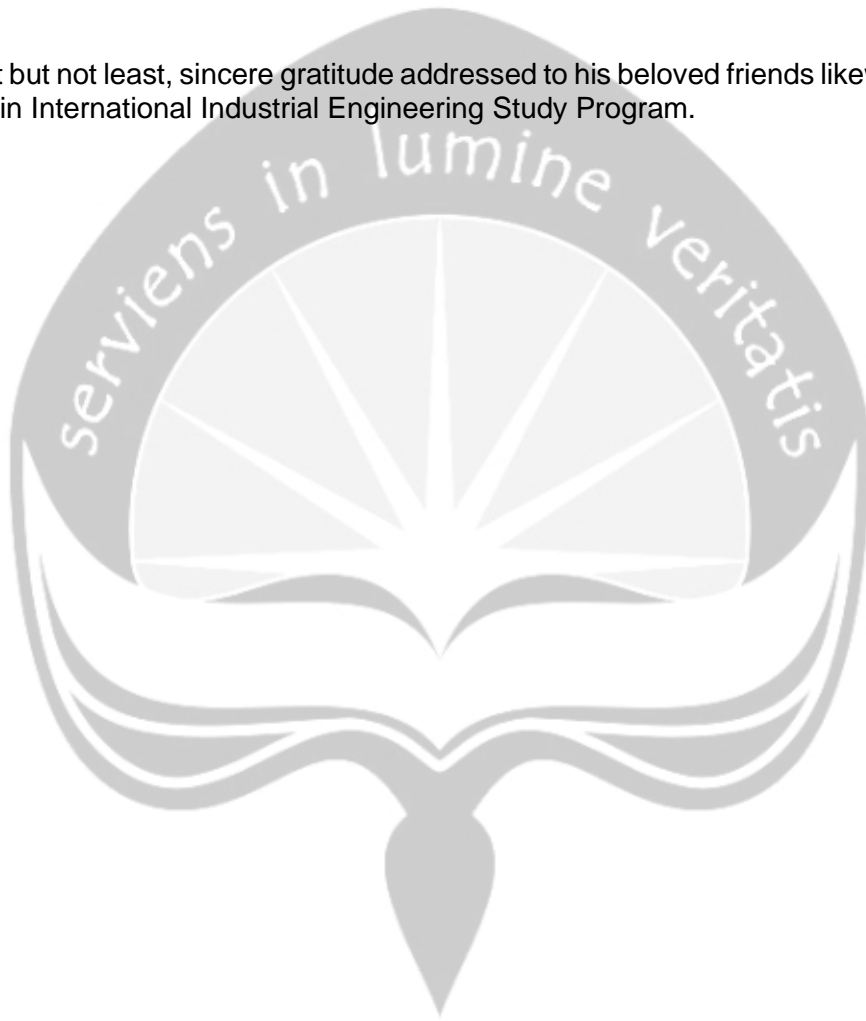


TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	Cover Page	i
	Identification Page	ii
	Declaration of Originality of Research	iii
	Acknowledgments	iv
	Table of Contents	v
	List of Tables	vii
	List of Figures	viii
	Abstract	x
1	Introduction	1
	1.1. Background	1
	1.2. Problem Formulation	3
	1.3. Research Objectives	3
	1.4. Scope and Limitation	3
2	Literature Review and Theoretical Background	4
	2.1. Literature Review	4
	2.2. Theoretical Background	9
3	Research Methodology	14
	3.1. Preliminary Stage	14
	3.2. Problem Formulation	14
	3.3. Data Collection	14
	3.4. Analysis and Discussion	15
	3.5. Conclusion	15

	3.6. Report Writing Stage	16
4	Data Collection and Company Profile	18
	4.1. Company Profile	18
	4.2. Data Collection	18
5	Data Processing	20
	5.1. Flowchart of Text Mining from Twitter API	21
6	Conclusion	52
	References	53



LIST OF TABLES

Table 2.1.	Literature Review Summary	6
Table 5.1.	Text Before Translation	24
Table 5.2.	Text After Translation	24
Table 5.3.	Text Before Cleaning	26
Table 5.4.	Text After Cleaning	26
Table 5.5.	Text Before Stemming	28
Table 5.6.	Text After Stemming	28
Table 5.7.	Text Before Stopwords Removal	29
Table 5.8.	Text After Stopwords Removal	30
Table 5.9.	Bag of Word	31
Table 5.10.	Summary of Bag of Words	31
Table 5.11.	Term Frequency-Inverse Document Frequency	32
Table 5.12.	TF-IDF Summary	32
Table 5.13.	TF-IDF Dataset (Top 100)	35
Table 5.14.	Bag of Words Dataset (Top 100)	37
Table 5.15.	Verbs Dataset	39
Table 5.16.	Nouns Dataset	41
Table 5.17.	Adjectives Dataset	42
Table 5.18.	Locations Dataset	43
Table 5.19.	Interjection Dataset	44
Table 5.20.	Time Dataset	44
Table 5.21.	Flavor Dataset	45
Table 5.22.	Cafe Dataset	50

LIST OF FIGURES

Figure 2.1.	Drew Conway's Venn Diagram of Data Science	9
Figure 2.2.	Data Science Lifecycle	10
Figure 2.3.	NLP Pipeline	13
Figure 3.1.	Research Methodology Flowchart	17
Figure 4.1.	Twitter Dataframe Information	19
Figure 5.1.	Interrelationship Diagram of The Problem	20
Figure 5.2.	Text Mining Flowchart	21
Figure 5.3.	Twitter API Access Key	22
Figure 5.4.	Scraping All Possible Tweets	22
Figure 5.5.	Scraping Indonesian Tweets	22
Figure 5.6.	Venn Diagram of All Possible Languages	23
Figure 5.7.	Function To Translate Other Languages	23
Figure 5.8.	Function To Clean The Text	25
Figure 5.9.	Sastrawi Stemmer	27
Figure 5.10.	Word-Net Lemmatizer	27
Figure 5.11.	Function To Remove Indonesian Stopwords	29
Figure 5.12.	Function To Remove English Stopwords	29
Figure 5.13.	Merge Dataset	30
Figure 5.14.	Actual Word Cloud	33
Figure 5.15.	Adjusted Word Cloud	33
Figure 5.16.	POS Value Count of TF-IDF	34
Figure 5.17.	POS Count Plot of TF-IDF	34
Figure 5.18.	POS Value Count of Bag of Words	36
Figure 5.19.	POS Count Plot of Bag of Words	36

Figure 5.20.	Brick Cafe	46
Figure 5.21.	Terrace Cafe	47
Figure 5.22.	Ideologi Cafe	47
Figure 5.23.	Chingu Cafe	48
Figure 5.24.	Loko Cafe	48
Figure 5.25.	Coklat Cafe	49
Figure 5.26.	Cangkir Cafe	49



ABSTRACT

Data is one of the important features for every business and organization. It helps the company to make decisions based on statistics and facts. In line with the fourth industrial revolution where industrial practices are combined with the latest smart technology, there also exist more and more data. Any incumbents that cannot reshape and rejuvenate their core business doubtlessly shall be perished.

PT. X is a printing company which is still in the process of restructuring their body so that their business can keep pace with the times. With the development of technology and e-book, the company started to face difficulties to maintain their core business in printing. Somehow the dispute on the company printing's sustainability started to emerge. The idea of diversifying the company's products and services also to offer them into a wider section of the existing market becomes a hot topic and idea within the company. Building a cafe or coworking space becomes one of the decisions. But in further planning, the company is facing doubt about this idea, because they do not have any empirical study yet to show the customer's or people's opinion about the cafe. This idea is important in order to create strategies and concepts later on for the company. Another issue that the company face is that they do not have a clear idea of what metrics are important, and which cafe to follow or imitate. In this case, data mining can help out the company to objectify their plan of building a cafe or coworking space by serving insights regarding customer's opinions through social media analysis.

Keywords: Data Mining, Natural Language Processing, Bag of Words, Term Frequency-Inverse Document Frequency, *Python*

CHAPTER 1

INTRODUCTION

1.1. Background

Schutt (2016) implies that data science, as it's practiced, is a blend of computer science skills, statistics, and domain knowledge. However, data science is not merely about software engineering, because after debugging and scripting, measurement of its metrics must be taken care of. It is not merely statistics as well, because after statistics carried out, it needs to be computerized for better treatment of dataset that cannot be performed using traditional statistics. After all, data science can be stated as the civil engineering of the data. It is serviceable and versatile when dealing with both structured and unstructured data to extract knowledge and meaningful insights.

Shortly in this very data-driven era, if a business decides something out of the blue and merely based on intuition, it may lead to trouble. At this very moment, data is everything. According to the research conducted by Forbes in 2018, every single day, people all around the world produced around 2.5 quintillion bytes of data. There is another fact that over the last 2 years alone, 90% of the data in the world was generated. From this fact, it is very important for a business to have a clear understanding of what had happened and what to do next with this enormous amount of data. Some companies can use historical data to understand more about customer preferences or conduct data mining from an open source to have a clear comprehension of market behavior also being able to gain a clear idea of what to do right afterward.

PT. X is a printing company. Even though the core business of the company is on printing, but this company also engaged in the trading fields such as books and accessories. Recently the company also run a packaging business and formal letter design that focus on B2B service.

Since the shift from the foundation into a corporation, PT. X is still managing and reshaping its corporate body to be able to compete with other companies. One of the steps taken to accelerate the shifting process is by opening Research and Development Department. The very basic purpose of this department's establishment was to develop the core and entire system of the company. In facing

the 4th industrial revolution, the company expects to rejuvenate the system, one of which is in data processing as a guide for corporate decision making.

Somehow, it only took about 2 years for this Department to go almost extinct. This happens because so far this Department only works to assist the job of IT Department without giving any significant impact to the company yet. This affects the reduction in the number of employees working in this division. At the moment, the Research and Development Department only has 1 personnel. So far, based on the dialogue with the Human Resource Department, they think that this decision gives no significant harm towards the company, yet no significant development as well. To think about in short-term business continuity, there might be no effect. But to put it in a longer and bigger picture, somehow the Human Resource professes their unsteadiness about the long-term sustainability of their business considering emerging competitors with core business like them offering services that even more modern and sophisticated.

The Human Resource mentioned that the company has a plan to create some cafe or co-working space. This is due to the large capacity and area of the field owned by the company. Actually, this renting space business is already running since decades ago. But so far they only rent out an empty field for an event. The very first purpose that the company want to create some cafe or co-working space is because of the emerging trend of flexible workspace which the company has (large space). The second thing is the concern about the sustainability of the core business in printing. The advent of technology and e-books made conventional books have to hardly fight within the market to still find a place for readers. This makes the company has an idea to diversify their products or services and exploring new markets. Thus, in this case, the company wants to offer products or services to a wider section of their existing market. Somehow, the company faces difficulties in determining people's or customer's opinions about cafe. The company may have some idea but it is based on intuition and notions only. In the later building process, PT. X also needs some role model to follow or imitate, either in terms of concepts, strategies, or core ideas. Some empirical study needs to be conducted here so that the decisions that are going to be made are evidence-based and data-driven.

1.2. Problem Formulation

Based on the background, the problems that can be formulated in this research are:

- a. How the company understands the market opinion about cafes?
- b. How the company determines the key metrics from popular cafes to establish a sustainable cafe?

1.3. Research Objectives

The objective of the research is to:

- a. Conducting text or opinion mining from social media to gain insights regarding the market's opinion about cafes.
- b. Modeling the data to find the most popular cafe and extracting information regarding their metrics of superiority.

1.4. Scope And Limitations

For the research to be focused on its main objectives, it is necessary to limit the problem. The limitations of the problem in this research are:

- a. The research is done under Twitter API (Application Programming Interface).
- b. The date of research is started from January 1st, 2019 to December 31, 2019.
- c. The location of the research is targeted at Yogyakarta.

CHAPTER 2

LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1. Literature Review

Liu, et al. (2018) conducted a research about comment on Chinese online shopping platform. The purpose of the research is to analyze customer's interests and behavior on different characteristics of the products.

Cripps, et al. (2020) conducted a research about the use of Twitter as an innovation in business markets. The purpose of the research is to investigate the use of Twitter in business as a medium for knowledge sharing and to crowdsource information to support innovation and enhance business relationships in the context of business-to-business (B2B) marketing.

Lim and Buntine (2014) conducted a research about product opinions from tweets by leveraging hashtags and sentiment lexicon. The purpose of the research is to aggregate or summarize the opinions of a product from Twitter.

Rekha, et al. (2018) conducted a research about sentiment analysis and opinion mining on Twitter. The purpose of the research is to examine the sentiments from natural language texts and addressing various opinions about the iPhone 6.

Kamthania, et al. (2018) conducted a research on some E-Commerce Business. The purpose of the research is to propose a model for formulating business strategies based on the users' interests and location. They used Principal Component Analysis (PCA) to reduce dimensionality followed by K-Mode Clustering Algorithm for segmentation.

Nugroho and Prihandoko (2018) conducted a research about the public sentiment of West Java Governor Election using Naive Bayes Classification. The purpose of this research is to find the right architecture to develop it into the application of twitter opinion mining to know public sentiments toward the election of the governor of west java.

Salloum, et al. (2018) conducted a research about text mining techniques for extracting information from research articles. The purpose of the research is to observe and compare different techniques regarding information extraction.

Bae, et al. (2017) conducted a research regarding twitter analysis about South Korea Presidential Election. The purpose of the research is to capture the political trend that occurred on Twitter.

Bach, et al (2020) conducted a research about Industry 4.0 job advertisements. The goal of this research is to develop a profile of Industry 4.0 job advertisements, using text mining on publicly available job advertisements, which are often used as a channel for collecting relevant information about the required knowledge and skills in rapid-changing industries.

Amarouche, et al (2015) conducted a research about product opinion mining. The purpose of the research is to present a new source that helps and leads the company to identify, analyze and manage the various risks associated with its business/products.

Wei, (2016) conducted a research regarding computational linguistic that concerned with the statistical or rule-based modeling of natural language from a computational perspective. The purpose of the research is to identify crossroads of information retrieval that occurs within the proposed text.

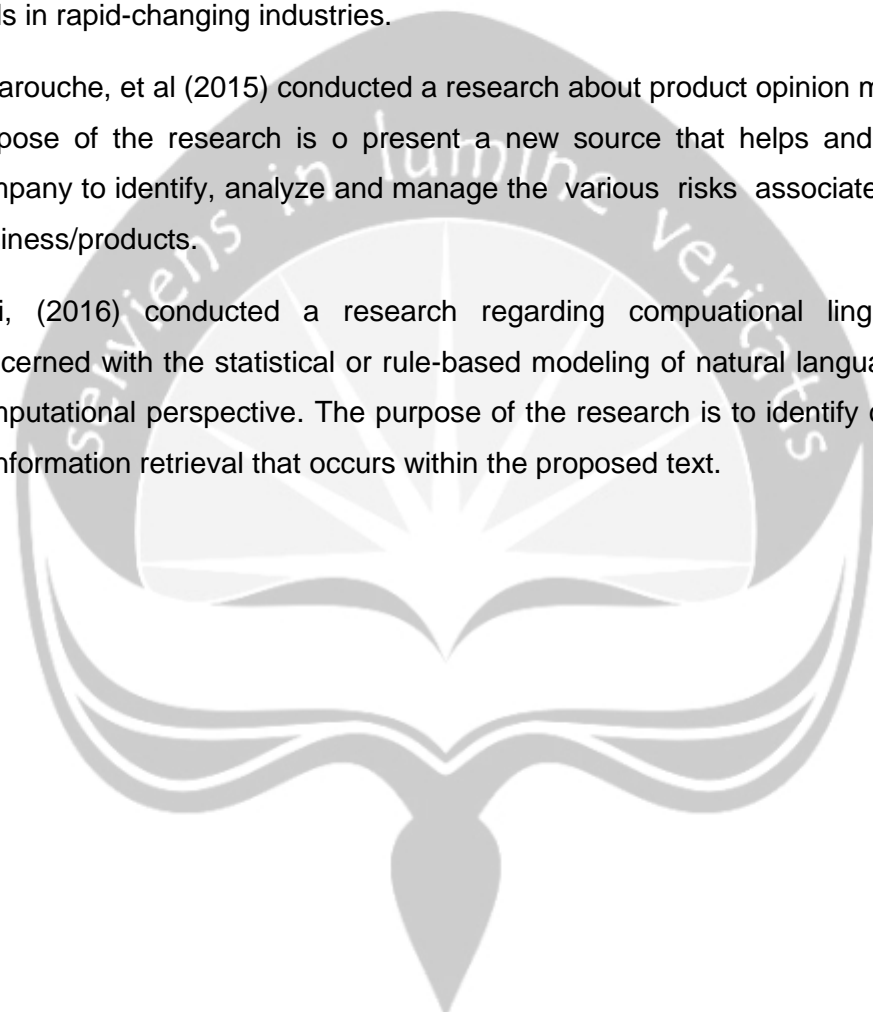


Table 2.1. Literature Review Summary

No	Researcher	Research Object	Method	Purposes
1	Liu, Z., Shen, Q., Ma, J., Dong, Z., (2018)	Online Shopping Platform	TF – IDF, SA-FCM, SA-PSO, Opinion Mining	To extract the comment targets in Chinese online shopping platform
2	Cripps, H., Singh, A., Mejttoft, T., Salo, J., (2018)	B2B Marketing across 5 Countries	Topic Modelling	To investigate the use of Twitter in business as a medium for knowledge sharing and enhance business relationships in the context of business-to-business (B2B) marketing.
3	Lim, K.W., and Buntine, W., (2014)	Product Opinions on Twitter	Opinion Mining, Sentiment Analysis, Topic Modelling, Sentiment Lexicon, Latent Dirichlet Allocation (LDA)	To aggregate or summarize the opinions of a product from Twitter
4	Rikha, P., Greetha, R., Karthika, S., (2018)	iPhone 6 Opinion on Twitter	Opinion Mining, Synset	To examine the sentiments from natural language texts and addressing various opinions about the iPhone 6.

Table 2.1. Literature Review Summary (Continued)

No	Researcher	Research Object	Method	Purposes
5	Nugroho, S., and Prihandoko (2018)	Public Sentiments of West Java Governor	Sentiment Analysis using Naive Bayes Classifier	To find the right architecture to develop it into the application of twitter opinion mining to know public sentiments toward the election of the governor of west java.
6	Salloum, S.A., Al-Emran, M., Monem, A.A., Shaalan, K., (2018)	Research Articles	Text Clustering, Association Rule Extraction, K-Means, Word Cloud	To observe and compare the result of information extraction using different techniques.
7	Bae, J., Son, J., Song, M., (2017)	Presidential Election in South Korea	Topic Modelling, Network Analysis, Sentiment Analysis	To predict the political trend during Presidential Election in South Korea.
8	Bach, M.P., Bertonce, T., Mesko, M., Krstic, Z., (2020)	Job Advertisement	Topic Mining, Clustering	To develop a profile of Industry 4.0 job advertisements on publicly available job advertisements

Table 2.1. Literature Review Summary (Continued)

No	Researcher	Research Object	Method	Purposes
9	Amarouche, K., Benbrahim, H., Kassou, I., (2015)	Product Opinion	Opinion Mining, Machine Learning, Natural Language Processing	To present a new source that helps and leads the company to identify, analyze and manage the various risks associated with its business or products.
10	Wei, W., (2016)	Computational Linguistics	Opinion Mining, Sentiment Ontology Tree	To identify crossroads of information retrieval that occurs within the proposed text.

2.2. Theoretical Background

Bruce (2017) explained that data science is a fusion of multiple disciplines, including statistics, computer science, information technology, and domain-specific fields. Meanwhile, according to Han (2017), data mining is a subfield of data science that focuses on extracting or “mining” knowledge from a large amount of Data. As a result, several different terms could be used to reference a given concept. To be known beforehand, data science is an emerging field in the industry which is continuously developing, and as yet, it is not well defined as an academic subject. To know the current landscape of what data science is, it is necessary to look online and see what ongoing discussions are taking place. Although it doesn’t explicitly tell the meaning, it at least tells how the world and professionals are perceiving it. In an updated discussion from Quora, here’s Meta market CEO Mike Driscoll’s answer:

“Data science, as it’s practiced, is a blend of Red-Bull-fueled hacking and espresso-inspired statistics. But data science is not merely hacking—because when hackers finish debugging their Bash one-liners and Pig scripts, few of them care about non-Euclidean distance metrics. And data science is not merely statistics, because when statisticians finish theorizing the perfect model, few could read a tab-delimited file into R if their job depended on it. Data science is the civil engineering of data. Its acolytes possess practical knowledge of tools and materials, coupled with a theoretical understanding of what’s possible.”

Driscoll then refers to Drew Conway’s Venn diagram of data science from 2010, shown in Figure 2.1.

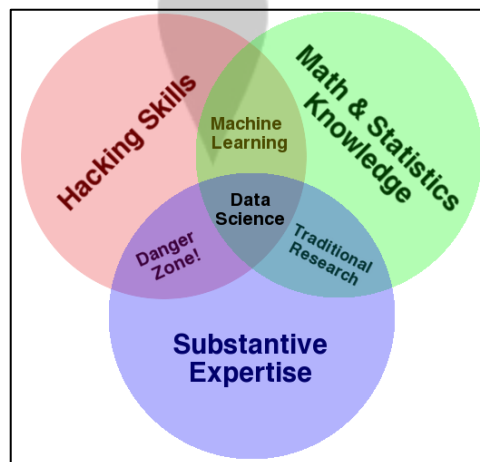


Figure 2.1. Drew Conway’s Venn Diagram of Data Science

2.2.1. Data Science Lifecycle

When working with a huge amount of data, it's always advantageous for a data scientist to perform and follow a well-defined workflow. According to Microsoft, there are at least five iterative lifecycle outlines of data science. This includes business understanding, data acquisition and understanding, modeling, deployment, and lastly customer acceptance.

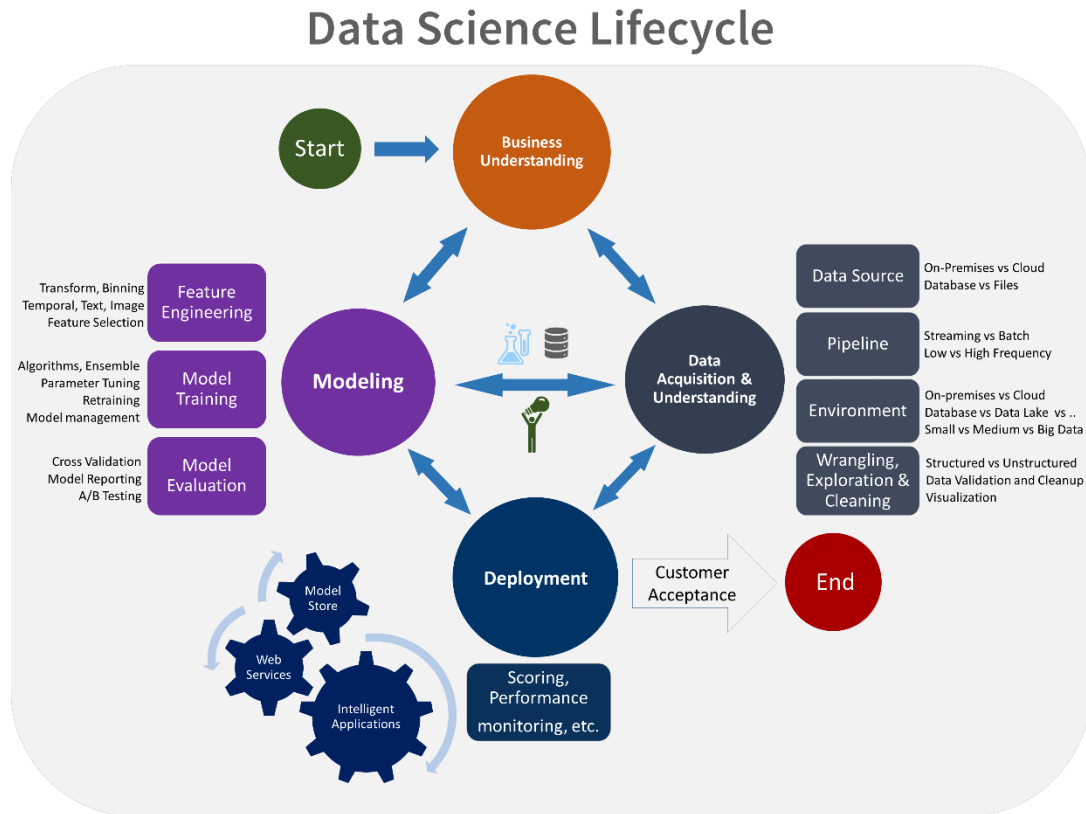


Figure 2.2. Data Science Lifecycle

2.2.2. Business Understanding

According to Géron (2017), one of the first important things data science can do is by framing the problem. Since building up the model is probably not the end goal if the why cannot be answered for the first time. Thus, asking the right question on the right problem become the role key beforehand. Critical thinking and questions like how the company expects to use and benefit from the model, what algorithm will be used, and what performance measurement will be used to evaluate the model are examples of the question to gain more knowledge regarding the problem and business understanding.

2.2.3. Data Acquisition

After gaining the right question over the right problem, then the data can be acquired for further understanding. In this step, it is called as data acquisition and understanding. This stage involves sourcing the right data, choosing the environment, cleaning the data, wrangling, and perform deeper exploration.

2.2.4. Data Cleaning

Most algorithms cannot work with missing values or features. Thus, in this stage, few functions are going to be performed. Generally speaking, there are three options in dealing with missing values:

- a. Drop the corresponding districts
- b. Get rid whole attribute
- c. Set the value equals to other value (mean, median, etc)

2.2.5. Exploratory Data Analysis (EDA)

O'Neil (2016) in her book quoting American Mathematician John Tukey that stated "Exploratory data analysis is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there". She added that exploratory data analysis is the first step towards building a model. Though it's traditionally presented as a bunch of visualizations such as pair plot, histograms, stem-and-leaf plots but EDA is critical of the data science process. The EDA is there to perceive the philosophy of the problem that data science tries to solve or might solve. Broadly speaking, basic tools of EDA are plots and graphs of summary statistics. It is one of the effective methods by plotting the distributions of all variables and try to see the pairwise relationship between variables using scatterplot matrices and generate a statistics summary for all of them.

2.2.6. Data Visualization

Cole Nussbaumer in his book, Story Telling With Data (2015) stated that data visualization is a technique used to communicate data or information by creating it in visual objects (for example; points, lines, bars, etc.) in a graph. A fundamental part of the data scientist's toolkit is data visualization. Although it is very easy to create visualizations, it's much harder to produce good ones. Part of the challenge

is that data visualization is a single step in the analytical process. There are two primary uses for data visualization which are to explore and communicate data. The two most popular types of data visualization are dashboard and infographic, both of which use a combination of graphics, text, and images to communicate data messages.

2.2.7. Natural Language Processing

Steven Bird on his book, *Natural Language Processing with Python* (2009) conveyed that natural means something that people use everyday, natural language defines kind of languages either English, Japanese, or Korean, while the term of Natural Language Processing (NLP) is where these set of data that contained languages are being processed together. In contrast to artificial languages such as programming languages and mathematical notations, natural languages have evolved as they pass from generation to generation, and are hard to pin down with explicit rules. The aim of NLP is “understanding” complete human utterances. The implementation of NLP is varied depending on the needs and purpose of the research. In-text mining generally people will use NLP for sentiment analysis, this sentiment analysis usually become very popular when there will be a presidential election in order to analyze people’s sentiment among one candidate to the other. Meanwhile, in market research people will dig in into opinion mining by understanding part of speech like verbs, nouns, adjectives contained in the words dataset to gain insights regarding their targeted customer or market’s behavior. Despite differences in usage among these purposes, most of the text mining will pass several processes, such as tokenization, stopwords removal, and stemming or lemmatization. Those three major processes are the basic and most important one to generate a result with good quality. Tokenization will be the initial step where data or sentences are being tokenized or splitted up into a single word index. Afterward, the tokenized data can be processed further by removing stopwords and mold them into its root word by stemming or lemmatization. Subsequently, the data can be processed furthermore depending on the needs or purpose of the research. If it is destined for sentiment analysis then the data can be matched up to the sentiment analyzer or compared with positive or negative text classifiers. If the use of the research is for determining the customer’s opinion, then the model can be continued to the model building of either Bag of Words or Term-Frequency Inverse Document Frequency.

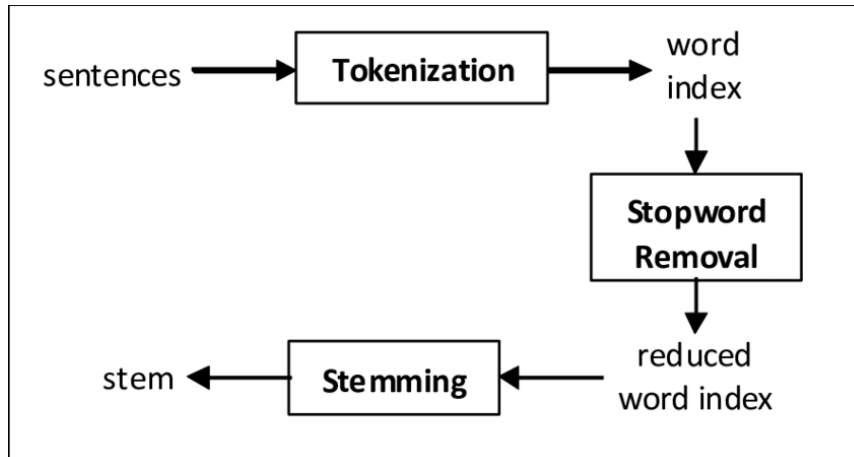


Figure 2.3. NLP Pipeline

2.2.8. Bag of Words (BoW)

Bronwlee (2017) mentions that a bag-of-words model (BoW), is a way of extracting features from the text for use in modeling, such as Machine Learning Algorithm. Bag-of-words is a representation of text which portrays word's occurrence within a document. Overall bag-of-words is containing of two major things which involve a vocabulary of known words, also a measure of the presence of known words.

2.2.9. Term Frequency – Inverse Document Frequency (TF-IDF)

A central question in text mining and natural language processing is how to quantify what a document is about, or in a short word is how a word's importance can represent the whole document. For example, there is a word: "She is beautiful". The word "she" and "is" will come often, and somehow the "beautiful" is one word that can represent this document. If this word changed into "ugly" than the meaning will be a whole different thing. Julia Silge and David Robinson in their book, Text Mining with R (2017) state that TF-IDF is intended to measure how important a word is to a document in a collection (or corpus) of documents. TF-IDF is calculated as follows, $TF(t) = (\text{Number of times term } t \text{ appears in a document}) / (\text{Total number of terms in the document})$, $IDF(t) = \ln(\text{Total number of documents} / \text{Number of documents with term } t \text{ in it})$.

CHAPTER 6

CONCLUSION

1. From the result of data mining Twitter API, note in advance that the research that has been conducted is not a mere justification or direct solution to build the cafe. Since the ultimate purpose is to answer the question regarding opinion mining, therefore from the result it can be concluded that most people spent their time in the cafe for a refreshing purpose justified by the word “*nongkrong*” and “refresh”. Most people regarding this activity are spending their time on the intention of social media content such as uploading or posting a photo. These action words are in line with the most popular cafe which possesses excellence in photography. The customer's favorite entertainments at the cafe are movies and music. When it comes to the cafe, the most popular location is placed on Sleman, around Selokan Mataram and Babarsari. It can be confirmed too that somehow price is still an issue for many people in Yogyakarta when they go to the cafe, justified by frequent discussions and words mentioning about “*menu*”, “*harga*”, “*murah*”, and “*porsi*”. About the time, it can be concluded that the most occupied time is on Weekend, specifically at night.
2. The result of most popular cafes is based on the most frequent words contained in the Bag of Words (BoW) model. From the elucidation of the top favorite cafe, the first and foremost suggestion towards the company will be to pick a specific niche. Since all of those most influential cafe got their superiority by serving special treatment towards a specific niche only. In order to do that, then the company should follow several metrics and considerations such as building an aesthetic space design, strategic placement of the location, also routine engagement or event towards the customer that belongs to the market niche.

REFERENCES

- Amarouche, K., Benbrahim, H., Kassou, I., (2015). Product Opinion Mining for Competitive Intelligence. Research Gate
- Bach, M.P., Bertoncel, T., Mesko, M., and Krstic, Z., (2020). Text Mining Of Industry 4.0 Job Advertisements. Research Gate
- Bae, J., Son, J., and Song, M., (2017). Analysis of Twitter for South Korea Presidential Election by Text Mining Techniques. Research Gate
- Bird, S., Klein, E., and Loper, E., (2009). Natural Language Processing with Python. Sebastopol, California: O'Reilly Media, Inc
- Browlee, J. (2017). Deep Learning For Natural Language Processing: Develop Deep Learning Models for Natural Language in Python.
- Bruce, P., & Andrew, B., (2017). Practical Statistics for Data Scientist. Sebastopol, California: O'Reilly Media, Inc
- Cady, F., (2017). The Data Science Handbook. USA: John Wiley & Sons, Inc.
- Cripps, H., Singh, A., Mejtoft, T., and Salo, J., (2018). The Use Of Twitter For Innovation In Business Markets. Emerald Insights
- Geron, A., (2017). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. California: O'Reilly Media, Inc
- Grus, J., (2015). Data Science From Scratch. Sebastopol, California: O'Reilly Media, Inc
- Han, J., & Micheline, K., Jian, P., (2017). Data Mining: Concepts Techniques. Morgan Kauffman, pp. 3-8
- Knaflic, C.N., (2015). Story Telling With Data: A Data Visualization Guide for Business Professionals. Hoboken, New Jersey: John Wiley & Sons, Inc.,
- Kuhn, M., & Johnson, K. (2018). Applied Predictive Modelling. New York: Springer Science and Business Media LLC
- Laudon, K.C., & Jane, P.L. (2007). Management Information System. Jakarta: Salemba Empat

- Lim, K.W., & Buntine, W., (2015). Twitter Opinion Topic Model _ Extracting Product Opinions from Tweets by Leveraging Hashtags and Sentiment Lexicon. Research Gate
- Liu, Z., Shen, Q., Ma, J., Dong, Z., (2018). Research On Comment Target Extracting In Chinese Online Shopping Platform. Emerald Insights
- McKinney, W., (2013). Python for Data Analysis. Sebastopol, California: O'Reilly Inc
- Nugroho, S., & Prihandoko (2018). Architecture of Text Mining Application in Analyzing Public Sentiments of West Java Governor Election using Naive Bayes Classification. Research Gate
- Peng, R.D., & Elizabeth, M., (2015). The Art of Data Science. Lean Publishing
- Rikha, P., Greetha, R., and Karthika, S., (2018). Twitter Opinion Mining and Boosting Using Sentiment Analysis. Research Gate
- Russel, R., (2018). Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python. CreateSpace Publishing Platform
- Salloum, S.A., Al-Emran, M., Monem, A.A., and Shaalan, K., (2018). Using Text Mining Techniques for Extracting Information from Research Articles. Research Gate
- Schutt, R., & Cathy, O., (2014). Doing Data Science. Sebastopol, California: O'Reilly Media, Inc
- Silge, J. & Robinson, D. (2017). Text Mining with R. Sebastopol, California, O'Reilly
- Wei, W., (2016). Analyzing Text Data for Opinion Mining. Springer Link