SENTIMENT ANALYSIS AND TOPIC MODELING

Netflix

TABLE OF CONTENTS

Problem statement

Knowing people's perceptions toward Netflix services and the most frequently words used when they express about their emotions. In addition to, the main topics used in the social media twitter network platform to ward Netflix. By conducting sentiment and frequency   analysis and topic modeling.

Abstract

After a decade of securing subscribers increasing, Netflix announced a subscriber loss. It also expects further losses from its subscriber base. The reason behind this decline is not clearly known. Netflix claimed that the reasons are password sharing, inflation, competition, and the Russian war. Other points of view said these reasons affected Netflix subscribers base, but it is not the only reason customer dissatisfaction is an additional reason. In this study we are answering two main questions; what is people's sentiment towards Netflix? And what are the main topics people discussing about Netflix on social media? From Twitter we extracted 14,000 tweets using RapidMiner, after the cleaning process it became 4,160 tweets. After conducting sentiment analysis, we found that 2,014 has positive perception of Netflix while 780 have negative sentiment, and 1,366 are neutral. The overall sentiment score is positive. Also, the most topics related to Netflix people talk about are Netflix, premium, spotifi.

1. Introduction

Knowing how people feel is a great power that helps in assessing current situations and making the right decisions. This is the main reason for the increasing interest and significance of sentiment analysis and text mining as they give insights into how people perceive a product, topic, or person. In addition to highlighting gain and pain points and determining the main topics people talk about.

Text mining is the process of getting meaningful worthy information from text for a specific objective [1]. The main difference between text mining and data mining is the type of the data; data mining works on structured data while text-mining deals with unstructured data (text) and transfers it to a structured form to identify patterns and get insights [2]. Text mining included under its umbrella sentiment analysis and topic modeling [3].

Sentiment Analysis is an approach of natural language processing (NLP) that reveals emotions in texts to define whether it is negative, positive, or neutral towards a specific topic, object, person, brand, or any other thing [4]. It applies text analysis and natural language processing to get significant subjective information from text [5]. Sentiment analysis consists of two main approaches: the machine approach learning and the lexicon approach. In the machine learning approach, a machine whether trained or not (supervised or unsupervised) detects the emotions in text. On the other hand, the lexicon approach in which lists of negative and positive words with negativity and positivity weights that lie between [-1:1] are used to determine emotions in text.

Topic Modeling is widely used in trends analysis, it is used to deduce hidden topics in a text by observing words and deriving hidden meanings to understand the overall trend [4]. It is an unsupervised machine learning model applied to unstructured data that identifies words and phrases with similar meanings and groups them into a topic, using statistical techniques to determine the topics included in text data. Topic classification is like topic modeling, the only difference is that topic classification is a supervised machine learning technique.

In this era in which people use the internet extensively, especially social media platforms to express their opinions and emotions, it becomes essential to know how they feel and think in many fields such as business and politics. Twitter is a social platform that has 229m daily active users and allows them to post tweets (messages with a maximum of 140 characters) [6]. Twitter restriction of characters makes users focused and concise about their feelings and opinions which makes it easier to analyze them. Also, Twitter provides Application Programming Interface (API) that allows analysts to retrieve real-time tweets to analyze from various users [7].

1.1 Research Problem

Netflix is the world's leading streaming service with 222 million subscribers in over 190 countries worldwide. Unexpectedly, for the first time in a decade, Netflix announced losing subscribers in its quarterly earnings report with further subscribers losing expectations. It is expected to lose around 2 million subscribers in the 2022 second quarter. Netflix blamed password sharing, inflation, competition, and the Russian war. Spencer Adam Neumann (Chief Financial Officer in Netflix) said Russia's invasion of Ukraine had indirect effects on Europe, MENA, and Africa. And these effects are clearly shown in Central and Eastern European countries. He also said that a little bit of competition contributes to this decline [8].

1.2 Research Objective

This research has two main objectives. The first objective is to determine people's perception and satisfaction with Netflix. Do they perceive it naturally, negatively, or positively? And what are the keywords most frequently used with Netflix which give insights about what makes them upset or happy? The second objective is to know the main topics that are discussed by people related to Netflix.

1.3 Research Questions

This research will answer the following questions:

Q1: What are people's perceptions of Netflix?

Q2: What are the keywords most frequently used with Netflix?

Q3: What are the main topics discussed by the people on social media related to Netflix?

1.4 Research Organization

This research contains six chapters, introduction, literature review, methodology, findings, discussion, and conclusion.

The introduction includes a brief about text mining, sentiment analysis, and topic modeling and why they are important. The literature review is a background about the research topic. In the methodology chapter, there is a description of how to extract and analyze people's tweets about Netflix on RapidMiner. In findings, declaring our analysis results. The discussion chapter includes our understanding of the findings and recommendations. And finally, the conclusion which summarizes our research outcomes.

2. Literature Review

In 1998 Reed Hastings and Marc Randolph established Netflix, the first website for renting and selling DVDs by mail. By 2006 Netflix membership grew to five million in the United States. In 2007, Netflix introduced streaming, it was a new model for entertainment. After that, it started to expand to other countries rather than the USA, and now it has subscribers in 190 countries around the world and covers about 62 languages [9]. This model's success reshaped the entertainment industry in the world.

The success of the Netflix streaming service encourages others to enter this industry such as Amazon Prime Video, HBO Max, Disney Plus, Hulu, and Peacock. The fierce competition among top streaming platforms over intellectual property pushes some of them to start their original production. According to Statista, Netflix released 129 original content titles all over the world in the 2021 third quarter [10].

During Covid -19 pandemic, unlike other industries, it helped the streaming industry to grow significantly. For a decade, Netflix didn't lose subscribers, and it was expected that it would gain about 2.5 million subscribers in 2022. Therefore, it was a shock when it announced subscribers losing with more losing expectations. Netflix points out that the Russian war is one of the main reasons for this loss, as it lost around 700,000 subscribers after suspending its services in Russia.

People in Netflix think that without this war they were supposed to gain subscribers as this war also affects the world economy and indirectly affects Europe especially Central and Eastern European countries, Africa, and MENA [8].

Password sharing also could be one of the problems that cause this decline. It is estimated that 100 million households around the world access Netflix services through password sharing which make it difficult to convert them to real subscribers. This is why Netflix started trying to put restricted controls on password sharing to turn some of them into subscribers. As a trial, Netflix started to ask subscribers to pay an additional $3 a month if they share their passwords with others outside their household in some countries in South America [8].

Some people say that the sharp increase in Netflix fees around the world could be a reason. For instance, UK subscribers are now paying tribble what they paid two years ago, but Netflix said that customers are satisfied with this increase as Netflix has the best retention in the industry and it remains significantly positive revenue.

Another point of view says that Netflix needs to care more about its service quality, the type of content it provides, and having more diversified content. As customer satisfaction could be one of the reasons that cause subscriptions. Some people think the huge production of Netflix has a lot of works that couldn't be considered real artwork, quantity over quality. What we try to do in this research is to discover if customer dissatisfaction is a hidden reason that contributed to this decline. And if so, identify what makes them unhappy with Netflix services by analyzing people's tweets about Netflix using RapidMiner.

3. Methodology

3.1 Data Structure

Our data was obtained in the shape of raw tweets from twitter social platform by using unstructured data method when collecting. Unstructured data refers to that the information obtained from this data has not been predefined structured, and that make it more difficult to get and understand results when using traditional software programs compared to use structured data. For our unstructured data we will use the tool called rapid miner program because this tool

organized and uses towards lexicon-based approach but do not support machine learning techniques.

Data collection, data can be collected using many ways and methods such as images, voices, video, open- end questions and social media platforms reviews. After that, data cleaning this phase after data collection phase and before sentiment analysis with text collected from twitter and its process to specify whether a part of text is positive, negative, or neutral.

3.2   Sentiment Approaches in RapidMiner

Rapid miner software is used to build up a web connection network. Also, it is social media mining stage that examines sentiment relationships in the combination of cloud mining procedures to indicate the produce. The stage of social media mining that verifies the medium of emotional relationships through cloud mining techniques with the aim of indicating the overall impact yield of thought states [11].

RapidMiner software can be used for both searching about our datasets from social network platforms and analytical platform that does not need programming and allows users to construct data analysis processes in a plug-and-play manner by connecting all operators together RapidMiner is a data science programmed, it makes users to do data analysis operations, also considered as a public source. It seeks to simplify the procedure by providing an easy-to-use GUI interface. RapidMiner is a data mining workflow tool for a variety of activities ranging from diverse data mining applications to different parameter optimization strategies [12].

3.3   Topic modeling in RapidMiner

Searching for the highly frequently topics used in the social media platform by users toward specific products or services. This process called topic modeling analysis by using RapidMiner tool. This type of analysis is done by using the Latent Dirichlet Allocation (LDA) method in RapidMiner. The process of topic modeling by using LDA method will

be discuss in this section. Before conducting topic modeling analysis, we need to make preprocessing process on the data collected to be suitable to make topic model analysis. LDA method is used to extract the most repeated topics from document to get valuable insights that might help business to improve [12].

3.4   Research Design

Nowadays, the new technology mostly appears in all sectors as well as the powerful of internet is already appears these two phenomena support the companies to conduct many types of analysis to know their users' expressions and perspective about their services. Customers express about their feeling and emotion through internet. That make it very supportive and easy to conduct. But before the technology appears peoples usually used to ask friends or family opinion and recommendation of products and services.

The report contain:

1.   Date Collecting
2.   Data Cleaning
3.   Sentiment analysis
4.   Frequency analysis
5.   Topic modeling

3.4.1 Data Collection

In our Netflix research, as shown in (**Fig.1**), we collected the raw twitter data up to 14,000 tweets, across three keywords we interested on it. As 10,000 tweets collected for Netflix, 2,000 tweets collected for Netflix helps and 2,000 tweets for Netflix movies. In this process we use three search twitter operators to get data contain all keywords we interested on it and save each keyword

data in separate file. Then we will combine these three files into one raw data file that will contain and present all search twitter Netflix data with all desired keywords, as shown in (**Fig.2**).
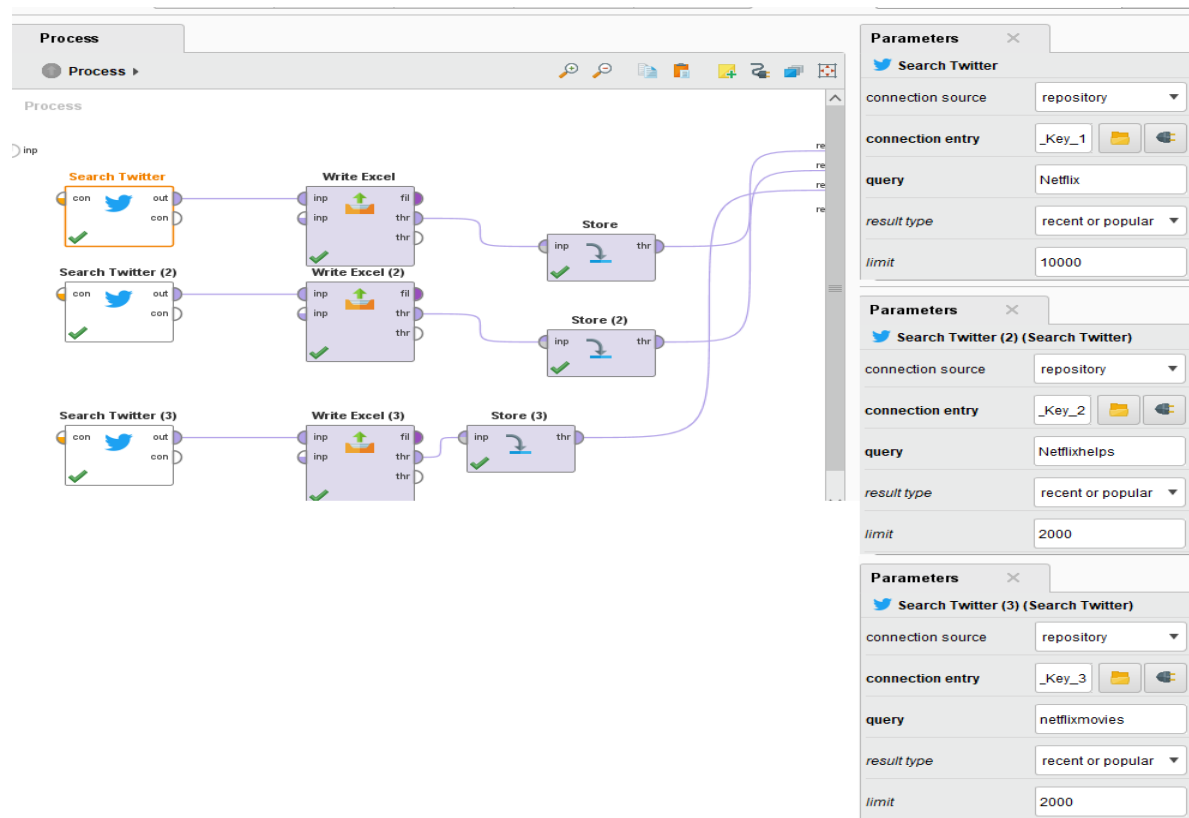

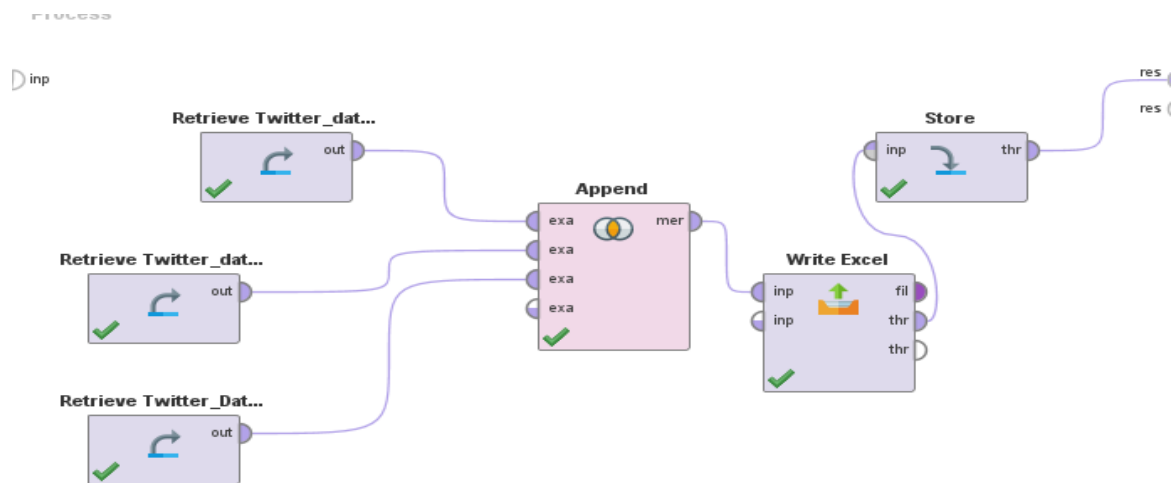**Fig.1.** explain the collection process of our data from twitter platform

Fig.2. shows the append process of twitter searching by keywords

9

3.4.2 Data Cleaning

Cleaning phase, after collecting the data from twitter social network platform, we need to clean the data due to the data collected is unstructured data and it will be more difficult to treat with it to get reasonable accurate information.
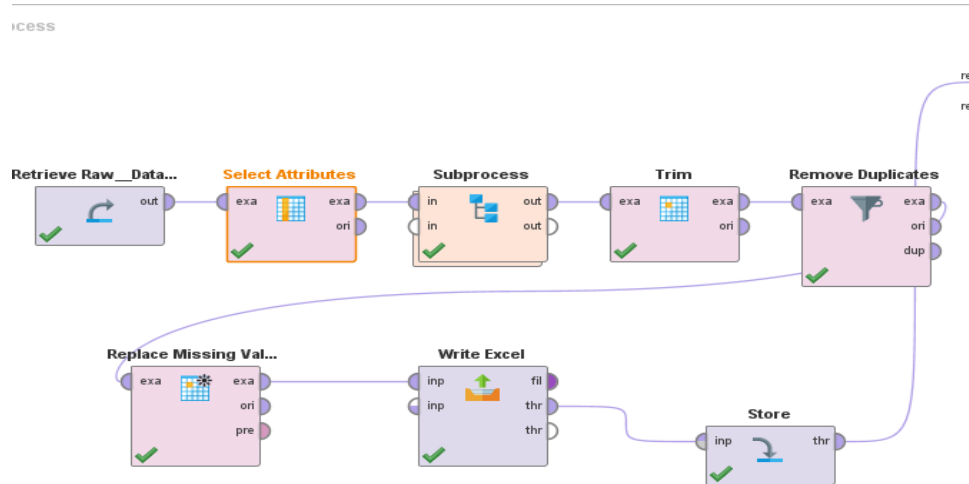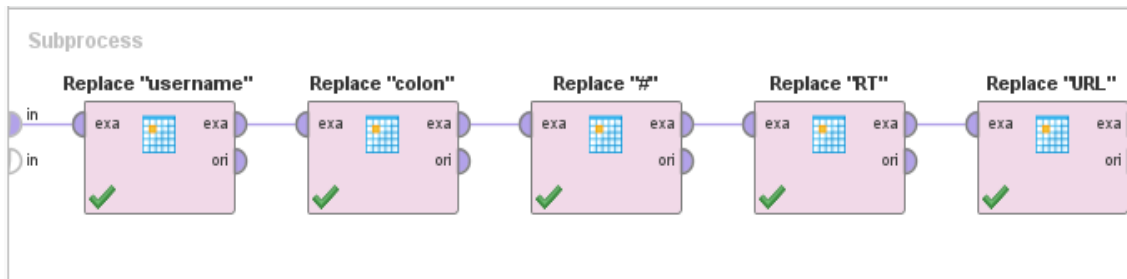

Fig.3. explain the cleaning process

Fig.4. explain the process of subprocess operator

The cleaning process is more important as shown in (**Fig.3**) and (**Fig.4**). Here the process will begin by retrieving the data collected, then we select some specific attributes

(columns) we need in our data with the main attribute "text", as we choose id, from user and created at attributes as shows in (**Fig.5**) by using "select attributes operator".

Then we make some cleaning process as shown in (**Fig.3**), that help us to make our data more accurate to be analyze. We will make five replacement process for some inappropriate characters in "text" attribute as following username, colon, hashtag, retweets, URL by using "subprocess operator".

After that as shown in (**Fig.3**), we use "trim operator" to deduct the white space in the beginning of raw twitter data for "text" attribute, by choosing "text" in its parameter, as shown in (**Fig.6**).

Also, as shown in (**Fig.3**), removing duplicate data from our "text" attribute is more important to get more accurate information from the analysis, we can do it by using "remove duplicates" operator.

On the other hand, and as shown in (**Fig.3**), "text" may contain missing values and that will disrupt our analysis. So, we will replace the missing value in "text" attribute by an appropriate character that will not affect the analysis using "replace missing values" operator.

Now, our data become suitable enough to make sentiment analysis on it, as in the following phase.
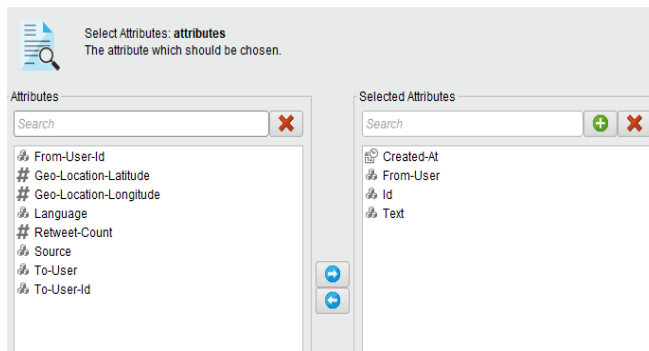

| Fig.5. selecting attributes in our dataset | Fig.6. remove the white space |

3.4.3 Sentiment Analysis

Sentiment analysis after data collected and cleaned that become more suitable for sentiment analyses to get some useful insights of our data about users' attitude and perspective toward our services whether to be positive, negative, or neutral reaction. As shown in (Fig.7) the sentiment analysis process.
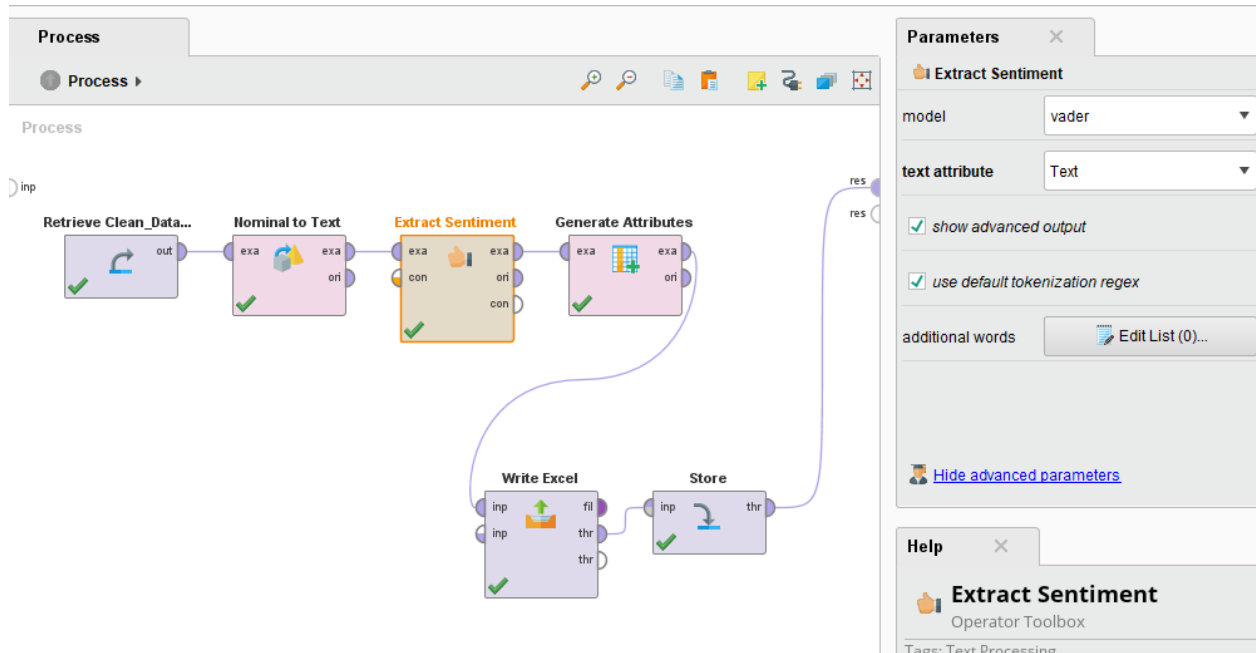
Fig.7. shows the sentiment analyses process

As shown in (Fig.7), we will begin with retrieving the cleaned data, then to make the analysis we should convert nominal data type to text data type due to the sentiment analysis analyzing text data so we will use "nominal to text" operator.

After converting data to text, that enabling sentiment analysis to be conducted by using "extract sentiment" operator. And VADER model for "text" attribute for the operator parameter to be selected. Using VADER as a dictionary-based approach to analysis with social networks text and work process whether a part of the text is positive, negative, or neutral. VADER is very sensitive to both emotions (positive/negative) and the strength (intensity) of feelings. In addition, it known as sentiment score that can be obtained by summing the strength (intensity) of each word in the text. It is more suitable when using twitter social network platform.

Fig.8. shows the function used to generate sentiment

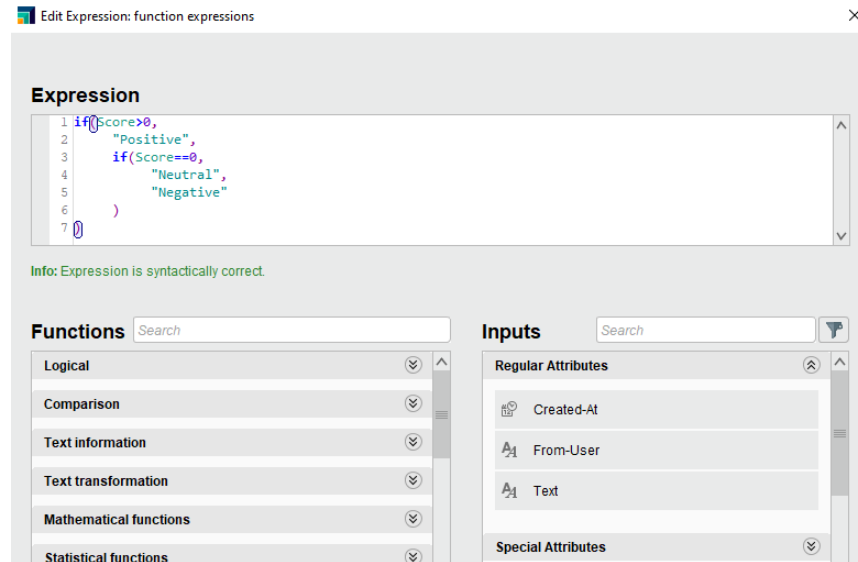As shown in (Fig.7), "Generate attribute" operator is used to generate the output of sentiment (positive\ negative\ neutral) based on the score by editing its function as shown in (Fig.8). This formula helps the analysis to be conducted as shown the expression "if", if the score > 0 that sentiment will be positive, while the score < 0 the result will be negative and if the score = 0 the result will be neutral. and then the sentiment output will be saved in excel sheet and store in RapidMiner repository.

Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At	Sentiment
1	1547294793...	-0.487	evil (-0.87) lik...	0.872	0.385	17	19	Resident Evil ...	Netflix	Jul 13, 2022 ...	Negative
2	1547649324...	0.513	dedicated (0....	0	0.513	9	10	In celebration...	Netflix	Jul 14, 2022 ...	Positive
3	1547288259...	1.385	thrilled (0.49) ...	0	1.385	36	39	We're thrilled ...	Satya Nadella	Jul 13, 2022 ...	Positive
4	1547764290...	0		0	0	7	7	Netflix gosta ...	lua	Jul 15, 2022 ...	Neutral
5	1547764288...	0.359	giving (0.36) ...	0	0.359	4	5	It's giving Har...	Cely	Jul 15, 2022 ...	Positive
6	1547764286...	1	clever (0.51) ...	0	1	25	27	What were th...	STRYKER of t...	Jul 15, 2022 ...	Positive
7	1547764283...	0		0	0	27	27	[INFO] Netflix ...	Queen Jhonny	Jul 15, 2022 ...	Neutral
8	1547764281...	0		0	0	3	3	Joseph em Y...	lua	Jul 15, 2022 ...	Neutral
9	1547764272...	-0.718	cruel (-0.72) ...	0.718	0	4	5	this is cruel, ...	i.s.r.a	Jul 15, 2022 ...	Negative
10	1547764269...	-0.128	want (0.08) f...	0.872	0.744	18	22	I don't want to...	Jesus Escala...	Jul 15, 2022 ...	Negative

Fig.9. sentiment analysis output for Netflix

After running the sentiment process, we get this output for Netflix. As shown in (Fig.9) sentiment output (positive\ negative\ neutral).

13

3.4.4 Frequency Analysis

Frequency analysis this type of analysis caries about the number of times a word appears in the document. In this analysis we split document into words and make some other process to get the highly frequency words appeared.

This process is important to know exactly how the customer feels about each word they Tweet.

We can make this analysis over all Netflix datasets as shown in (Fig.10), as well as we can conduct it for each sentiment output whether (positive\ negative\neutral) to know the words that mostly frequently appeared in the positive document same for negative and neutral.
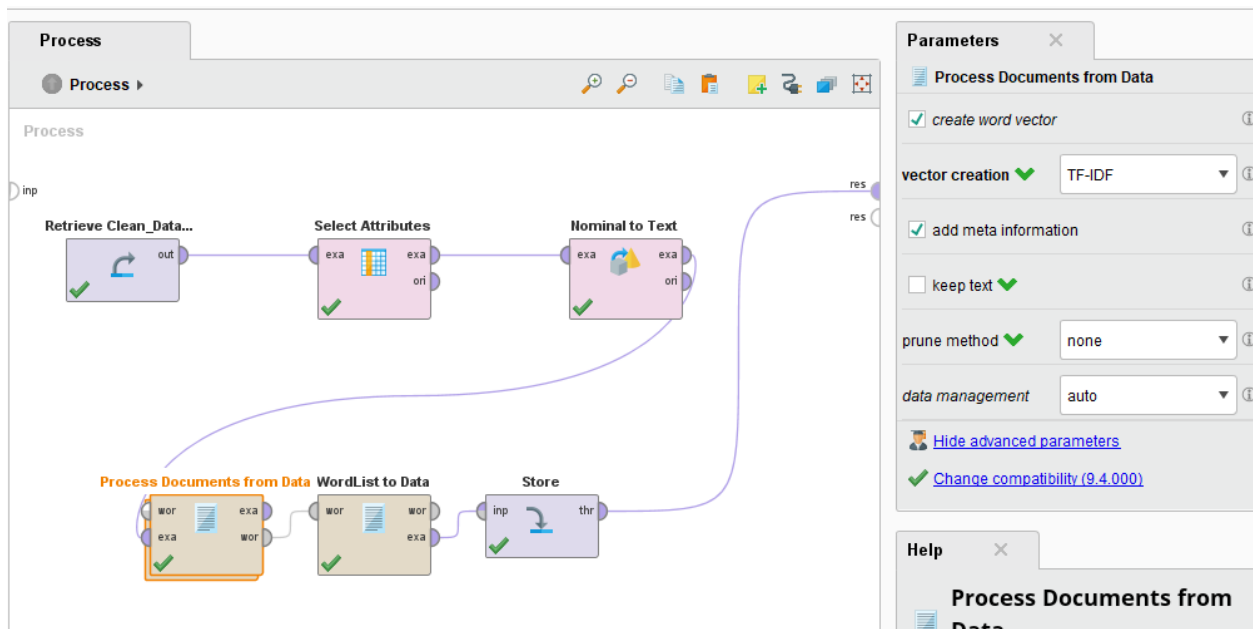

Fig.10. explain the frequency analysis process

Making frequency analysis over the whole Netflix data as (Fig.10) shows, requires firstly to retrieve the cleaned data, then select the "text" attribute due to the analysis will be conducted on it by using" select attributes" operator.

After that and as shown in (Fig.10), we will convert the selected nominal data attribute to string (text) data type, and it is very important not to miss any nominal value because it will be missing when convert it to text.
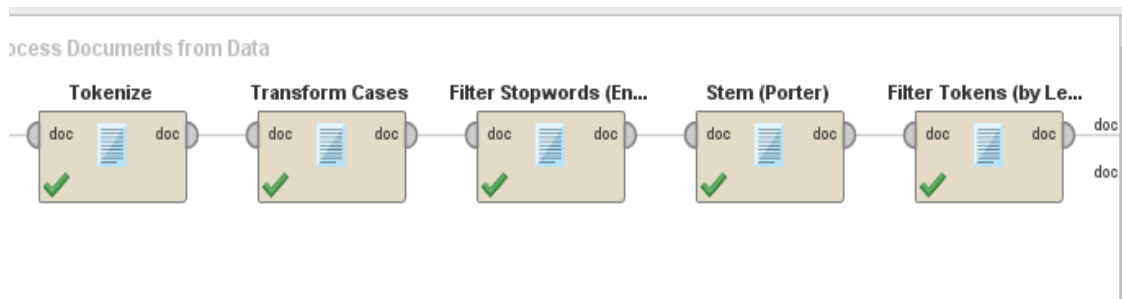

Fig.11. explain the process in the "process document from data" operator

After converting nominal to text and as we explain in (Fig.11), then we need to make some other process on the document as shown in (Fig.11) as following:

The "tokenize" operator is used to split every document or a few words into words or sequence of tokens.

The "transform cases" operator is used to transform all the upper cases letters to lower case, that standardize word to lower case.

The "Filter Stop words (English)" operator is used to remove the English stop words that not useful for our data analysis such as the, is, and from the document.

The "stem(porter)" operator to reduce the length of word until a minimum length is reach. As we set minimum character as 3 and maximum as 25 characters to avoid inappropriate character that will not be useful for analysis and may disrupt it these characters like as ', s.

The "Filter Tokens (By Length)" operator is used to filter tokens based on their length.

As shown in (Fig.10), add "wordlist to data" operator to convert created word list into a data set. Then store the dataset in the RapidMiner repository.

As we mentioned above that we can make frequency analysis for each sentiment outputs whether (positive\ negative\neutral). We can make this using the same process with slightly different in some specific parameters and filters of operators.

As we shown in (Fig.12), this process is applied for the frequency analysis for sentiment outputs whether (positive\ negative\neutral) the only different thing related to the operator called "filter examples" by clicking on it and make the choose to depend on desired output. If we need to apply this analysis and generate word list for positive sentiment, we will make the parameter filter by score > 0 if we want to make it for negative filter to make score < 0 and for neutral filter by making score equal to zero as shown in (Fig.13).
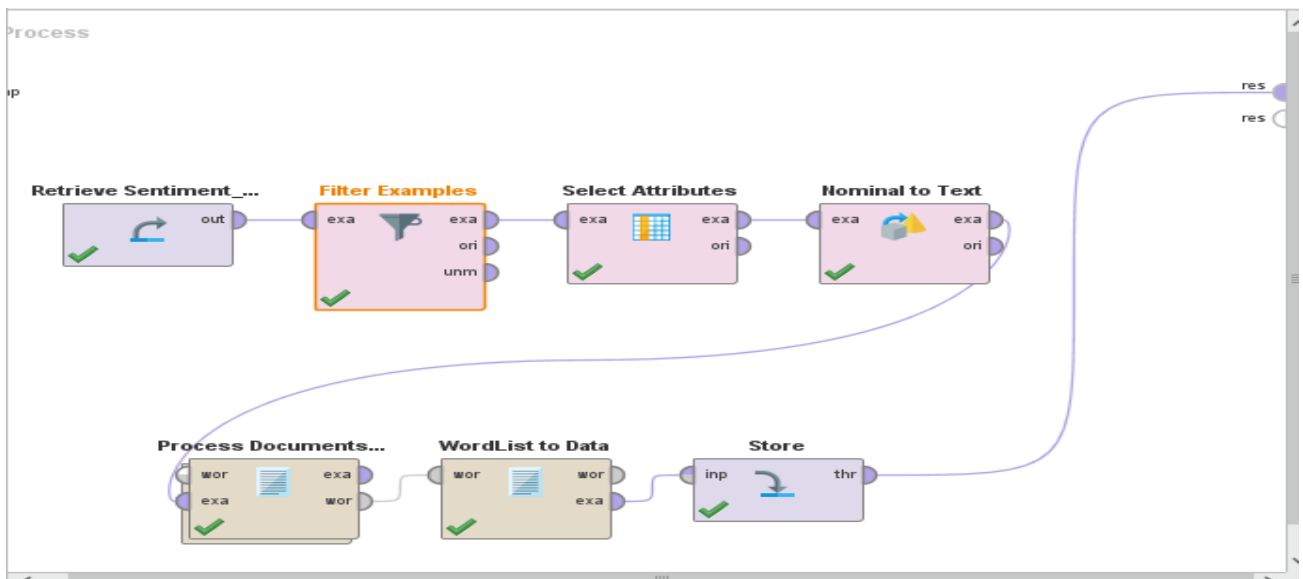
Fig.12. process of sentiment output frequency analysis

As shown in (Fig.12), the process of sentiment output frequency analysis begins with retrieve the sentiment data, then use "filter examples" operator to generate word list about the desired output. So, if we want to generate positive word list, we will make the filter by score greater than zero. On the other hand, if we want to generate negative word list, we will make the filter score smaller than zero. In addition to get the neutral word list, the score will equal to zero, as shown in the following (Fig.13).

Fig.13. explain different parameter for "filter example" operators

Then we choose the selected "text" attribute because the analysis will be conduct on the text.

Then we will use "nominal to text" operator to convert the selected nominal data attribute to string (text) data type.

Then conducted some process by using the operator called 'process document to data' operator, as shown in (Fig.12) as following same as (Fig.11):
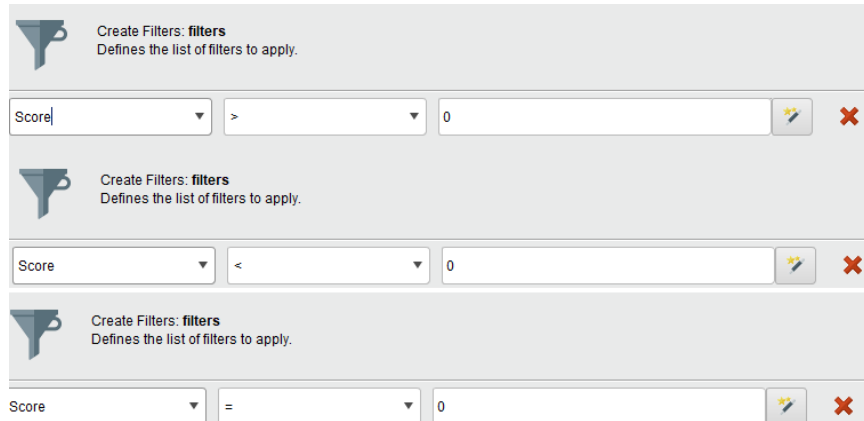
The "tokenize" operator is used to split every document or a few words into words or sequence of tokens.

The "transform cases" operator is used to transform all the upper cases letters to lower case, that standardize word to lower case.

The "Filter Stop words (English)" operator is used to remove the English stop words that not useful for our data analysis such as the, is, and from the document.

The "stem(porter)" operator to reduce the length of word until a minimum length is reach. As we set minimum character as 3 and maximum as 25 characters to avoid inappropriate character that will not be useful for analysis and may disrupt it these characters like as ', s.

The "Filter Tokens (By Length)" operator is used to filter tokens based on their length. As shown in (Fig.12), add "wordlist to data" operator to convert created word list into a data set. Then store the dataset in the RapidMiner repository.

17

3.4.5 Topic Modeling

Topic modeling, while conducting on our Netflix project using RapidMiner tool, the method will use called Latent Dirichlet Allocation (LDA), which is a method for analyzing a large set of documents.

As shown in (Fig.14), we retrieve the data after cleaning to enabling the topic modeling to be conducting, not only that but we make another preprocessing technique such "selecting attributes", "nominal to text", "data to document" and "loop collection". Then add the operator that conducting topic modeling analysis called Extract Topics from Documents (LDA) is used to categorize text in the document to a specific topic.
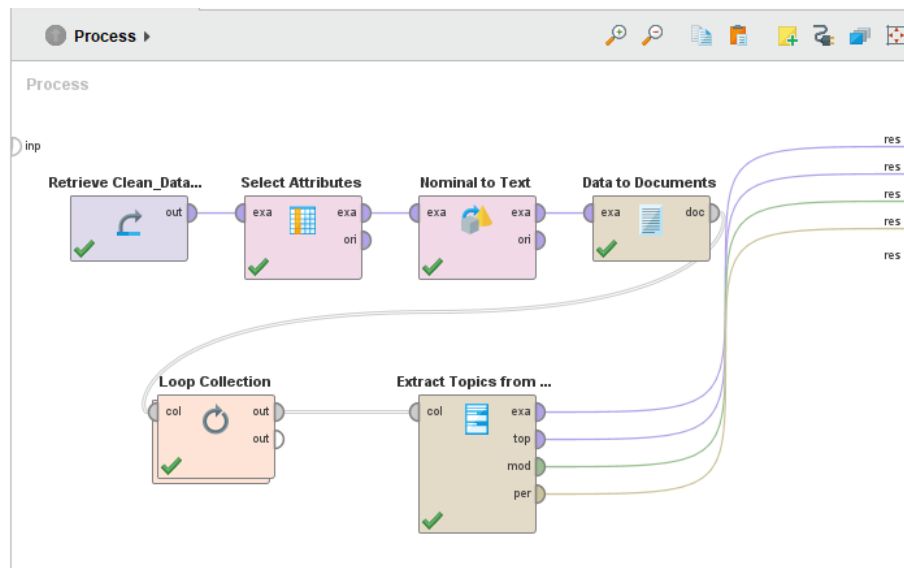

Fig.14. Top Modeling process

Firstly, we need to retrieve cleaned data, then we selected "text" attribute by using "select attributes" operator, then use "nominal to text" operator to convert the selected attribute from nominal data type to text (string) data type to conduct the analysis.

Then we use "data to document" to transform the clean Netflix dataset to a collection of documents by gathering a document for each example of the dataset

Adding "loop collection" operator to conduct a few subprocess, as shown in the following (fig.15):
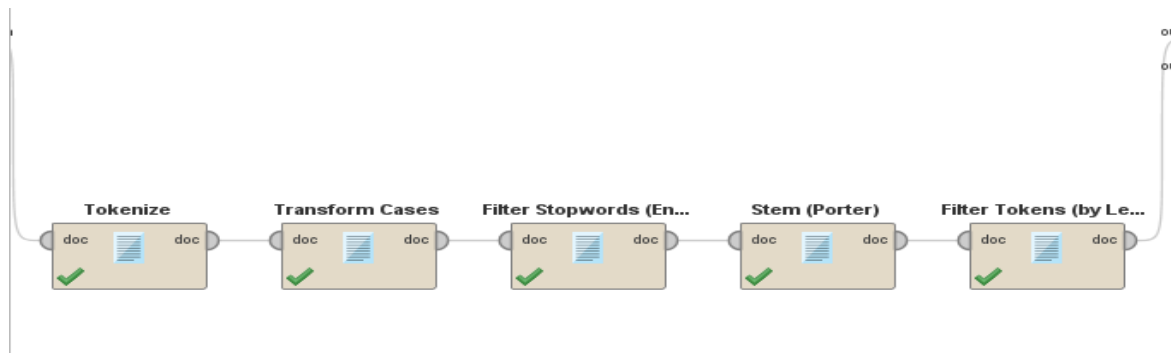
Fig.15. process including in the loop collection operator

The "tokenize" operator is used to split every document or a few words into words or sequence of tokens.

The "transform cases" operator is used to transform all the upper cases letters to lower case, that standardize word to lower case.

The "Filter Stop words (English)" operator is used to remove the English stop words that not useful for our data analysis such as the, is, and from the document.

The "stem(porter)" operator to reduce the length of word until a minimum length is reach. As we set minimum character as 3 and maximum as 25 characters to avoid inappropriate character that will not be useful for analysis and may disrupt it these characters like as ', s.

Then use "Filter Tokens (By Length)" operator is used to filter tokens based on their length.

From (Fig.14), Finally, add Extract Topics from Documents (LDA) to the process and configure the connection accordingly. LDA method allows you to identify topics in documents.

Run the process and we will get the result of topic modeling; we will show the output in the result and dissuasion section.

19

4. Finding

In this section we will discuss the finding from our Netflix twitter search. We tried to gather up to 14,000 raw twitter data containing the three main keywords, and after cleaning the data and remove duplicates tweets the raw twitter data become 4,160 tweets.

Sentiment analysis, after conducting the sentiment analysis we get the following general overview about the user's perspective toward our services and the words used in their expressions. As (Fig.16) shows the customers perspective toward Netflix services.
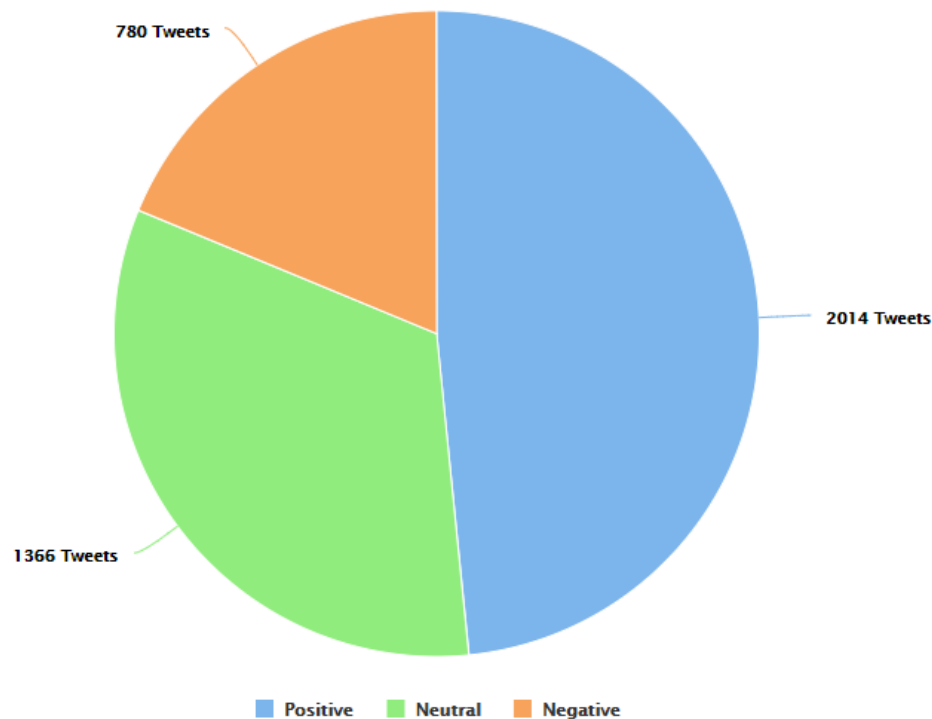
Fig.16. sentiment analysis

The above chart shows the twitter users perspectives toward the Netflix services. As shown from (Fig.16) that the most peoples reacted positively to Netflix.

And from the total of 4,160 tweets that complete the sentiment analysis for Netflix. The highest sentiment toward Netflix is positive with 2,014 tweets that mean it present 48.4% from the total tweets. On the other hand, having a negative sentiment tweets up to 780 tweets and that present 18.75% from the total while having the lower sentiment toward neutral by 1,366 tweets that refer to 32.8% of the total raw data.

ExampleSet

share   disnei   thank
account   movi
canva   get
show
netflix
youtub   readi
order   viu
trust   pleas   jual
spotifi   premium   watch
hbo

Fig.17. positive sentiment word list

As shown in (Fig.17), the word cloud analysis for positive sentiment. From this analysis we can know the most frequently words that used than other words when users reacted positively. The first five words that used heavily toward positive preceptive "Netflix", "watch", "premium", "share", "Spotifi". the biggest the word the higher usage of it in the twitter social platform.

Fig.18. negative sentiment analysis word list

As shown in (Fig.18), the word cloud for negative sentiment analysis. From this analysis we can get the most frequently words that used than other 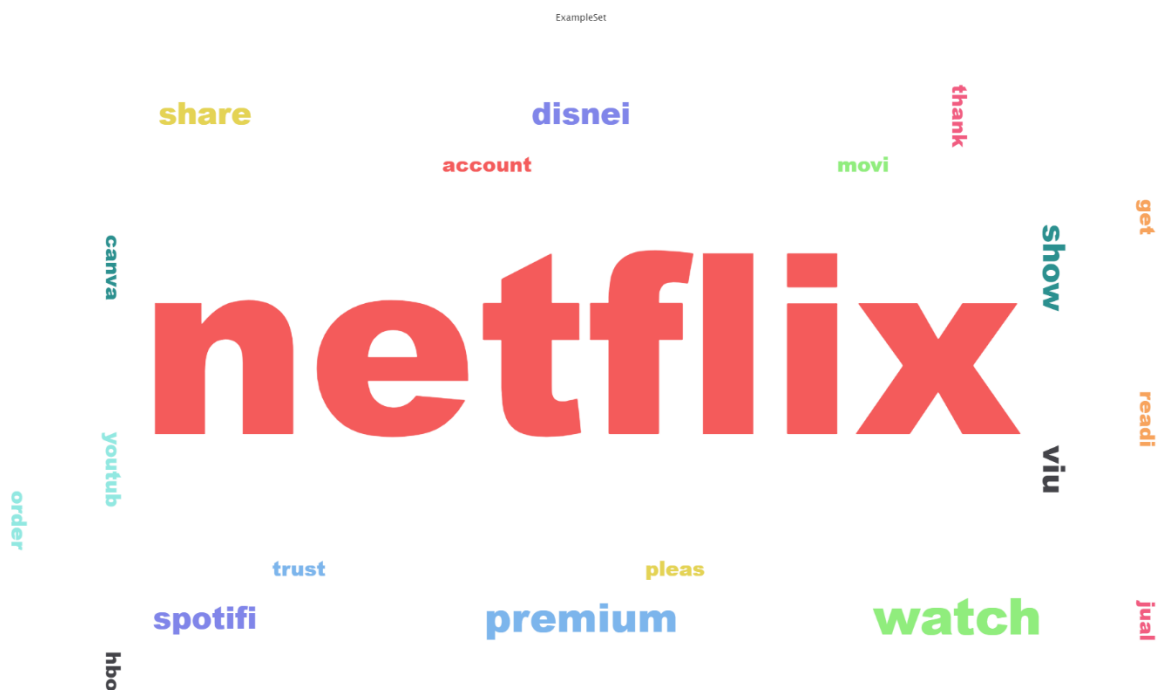words when users reacted negatively. The first five words that used heavily toward negative preceptive "Netflix", "watch", "premium", "Spotifi", "show". the biggest the word the higher usage of it in the twitter social platform when user react by negative emotions.

Positive Sentiment			Negative Sentiment		
word	total occurance	total in doucument	word	total occurance	total in doucument
netflix	1469	1323	netflix	3113	2871
watch	299	267	watch	571	508
premium	274	238	premium	478	392
share	246	193	spotifi	375	348
spotifi	222	204	show	325	287
show	219	189	youtube	289	284
viu	188	166	disnei	273	263
account	188	133	account	339	262
disnei	180	176	viu	283	261

Table.1 explain word list for positive and negative sentiment

22

As shown above in (Table.1), the words generate for Netflix sentiment analysis. For positive sentiment the Netflix word itself occurred in most frequently way totally by 1,469 times in 1,323 documents, follow up by watch and premium. Other many words can be used by users toward their positivity for service provide.

On the other hand, for word generate for Netflix negative sentiment. The word Netflix has the highest occurrence by 3,113 times in 2,871, follow up by watch and premium. Many other words be used when users' perception be negative.

Doucument \ Topic	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
0	0.001	0.004	0.003	0.003	0.218	0.763	0.001	0.003	0.001	0.003
1	0.003	0.012	0.011	0.005	0.002	0.004	0.003	0.007	0.002	0.95
2	0.002	0.006	0.004	0.79	0.001	0.005	0.001	0.002	0.001	0.189
3	0.01	0.044	0.032	0.018	0.004	0.844	0.008	0.015	0.004	0.022
4	0.004	0.029	0.023	0.008	0.004	0.031	0.005	0.012	0.004	0.88

Table.2 topic modeling output (LDA)

As shown above in (Table.2), the output of topic modeling analysis using Latent Dirichlet Allocation (LDA).

This table represent part of LDA output, while using this method the documents of our dataset grouped into one of the 10 topics which are from topic 0 to topic 9, even if they cover a variety of topics.

As previously stated, we use LDA analysis to count how many times the topic is repeated in the documents. From table (Table.2), we can discover what word has the highest weight from the list of words that we extract from tweets on Twitter using RapidMiner studio to make our sentiment analysis to help us know people's feelings. Referring to (Table.2) that represent the confidence level of each topic toward the document. As, document 0 has a high confidence value for topic 5 by 0.763, document 1 has a high confidence value for topic 9 by 0.95, document 2 has a high confidence value for topic 3 by 0.79, and document 3 has a high confidence value for topic 5 by 0.844, document 4 has a high confidence value for topic 9 by 0.88. Document 0 and Document 2 are related to Netflix, which is facing the problem of subscriber losses. This problem is still a

problem for Netflix, whose name is "Netflix." Document 1 and Document 4 show what people are thinking about Netflix, which the word is "Netflix". Finally, topic 3 is attached to Netflix's production to know which types of subscribers love the word "Netflix."

5. Discussion

Netflix's losing subscribers is a hot topic now. So, we made sentiment analysis from Twitter to find people's opinions. In this section of the research, we will explain our analysis from the data extracted from twitter social platform using RapidMiner tool. Based on the data collected, although Netflix is losing subscribers, people's opinions on Twitter are still positive, as it is obvious in Fig.16.

We have three questions that we will answer in this section:

Q1: What are people's perceptions of Netflix?
Q2: What are the keywords most frequently used with Netflix?
Q3: What are the main topics discussed by the people on social media related to Netflix?

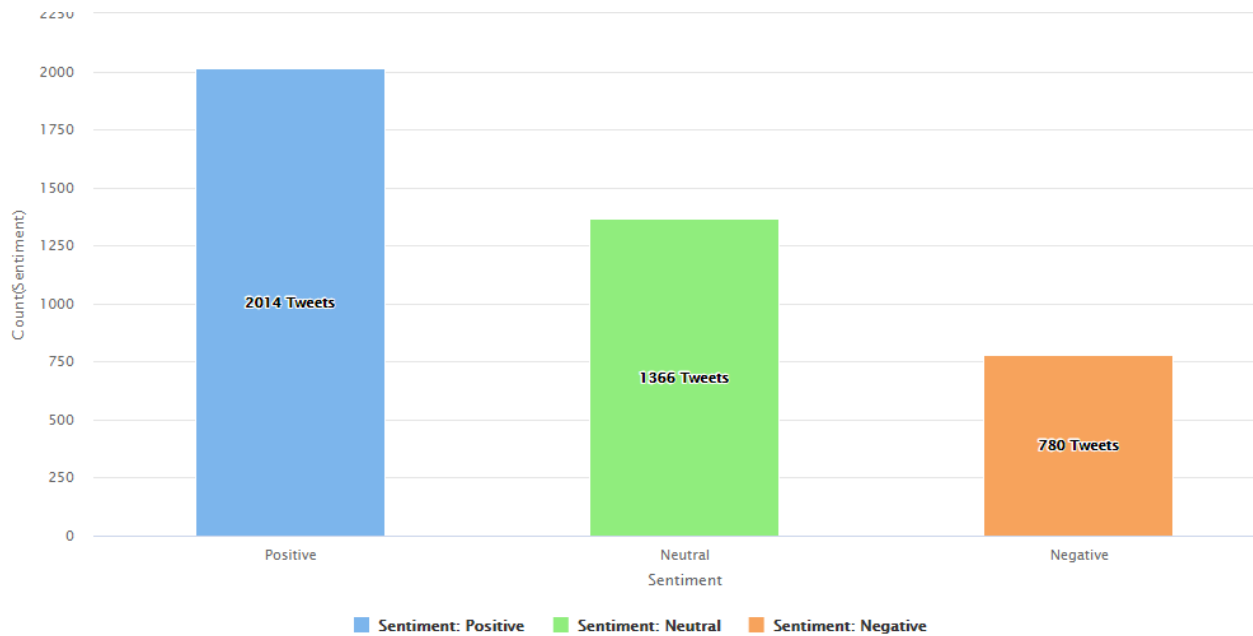## Q1: What are people's perceptions of Netflix?

Fig.19. people's perceptions sentiment analysis

From Figure.19, we see the viewpoint of Twitter users in general through their tweets about Netflix. The analysis shows that most of the tweets with a high percentage are positive, and after that, neutral come in second place in the highest tweets, and the lowest percentage of negative tweets. Positive is the highest because it has comments like "invited", "love" and "enjoying". Also, most of the subscribers recommend movies and series to new subscribers, like "Watch Mean Girl, It's Interesting." This comment has a 5 star, and "recommended to watch this spy x family." This comment has a 4.5 star. The neutral tweets rank second because they are general, and the subscriber does not use expressions such as "Where is Netflix documentary" and "Operation Romeo on Netflix was 10/10." negative because negative tweets like positive tweets speak about the production of Netflix's "strange things" and "just learn this series is not on Netflix."

Q2: What are the keywords most frequently used with Netflix?

Positive Sentiment			Negative Sentiment		
word	total occurance	total in doucument	word	total occurance	total in doucument
netflix	1469	1323	netflix	3113	2871
watch	299	267	watch	571	508
premium	274	238	premium	478	392
share	246	193	spotifi	375	348
spotifi	222	204	show	325	287
show	219	189	youtube	289	284
viu	188	166	disnei	273	263
account	188	133	account	339	262
disnei	180	176	viu	283	261

Table.3 explain word list for positive and negative sentiment

Unlike the previous section, in this section, we will explain the top ten keywords in more detail, both on the negative and positive sides. It is necessary to discuss the top ten keywords that are highly frequently appeared as results in this paper. In (Table.3), it is obvious that the Netflix word appeared on both the positive and negative side. Although it has the highest occurrence on both sides, the occurrence on the negative side is higher than the positive side. Netflix was highest in the negative because most of the users were not satisfied with Netflix's production or didn't find what they were looking for on the platform. It appears on the positive side because some subscribers recommend specific productions to new subscribers or love the production.

After this, it comes to watch and premium in the second and third places, but they are higher on the negative side than the positive side because most of the tweets are "I didn't love the witch hunter I watched" or "someone hacked my premium account".

Q3: What are the main topics discussed by the people on social media related to Netflix?

Row No.	topicla	word	weight
6	1	netflix	784
11	2	netflix	517
41	8	bulan	401
16	3	netflix	329
7	1	watch	318
46	9	netflix	268
1	0	month	259
31	6	netflix	250
42	8	netflix	246
2	0	netflix	231
26	5	netflix	196
47	9	thing	185
21	4	premium	184
32	6	viu	178

Doucument \ Topic	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
0	0.001	0.004	0.003	0.003	0.218	0.763	0.001	0.003	0.001	0.003
1	0.003	0.012	0.011	0.005	0.002	0.004	0.003	0.007	0.002	0.95
2	0.002	0.006	0.004	0.79	0.001	0.005	0.001	0.002	0.001	0.189
3	0.01	0.044	0.032	0.018	0.004	0.844	0.008	0.015	0.004	0.022
4	0.004	0.029	0.023	0.008	0.004	0.031	0.005	0.012	0.004	0.88

Table.4 main topics

As previously stated, we use LDA analysis to count how many times the topic is repeated in the documents. From (Table.4), we can discover what word has the highest weight from the list of words that we extract from tweets on Twitter using RapidMiner studio to make our topic modeling to help us know people's feelings. Referring to (Table.4) that represent the confidence level of each topic toward the document. As, document 0 has a high confidence value for topic 5 by 0.763, document 1 has a high confidence value for topic 9 by 0.95, document 2 has a high confidence value for topic 3 by 0.79, and document 3 has a high confidence value for topic 5 by 0.844, document 4 has a high confidence value for topic 9 by 0.88. Document 0 and Document 2 are related to Netflix, which is facing the problem of subscriber losses. This problem is still a problem for Netflix, whose name is "Netflix." Document 1 and Document 4 show what people are thinking about Netflix, which the word is "Netflix". Finally, (Table.4) is attached to Netflix's production to know which types of subscribers love the word "Netflix".

6. Conclusion

For this project, we used RapidMiner to extract data from Twitter. We analyzed this data using sentiment analysis. It is very useful in the analysis of any type of data extracted from any social media platform. Is a technique used in natural language processing (NLP) to determine whether data is positive, negative, or neutral? It is frequently performed on text data to help companies monitor customer sentiment and understand customer needs. The tools are essential for discovering and understanding customer sentiment. Companies that use these tools to understand how their customers feel can use them to improve their customers' experience. The tools provide insights into how companies can improve customer experience and service. It is a broad field that includes both supervised and unsupervised methods.

The purpose of this study is to discover people's opinions on Netflix from Twitter, especially after the loss of subscribers. To know people's opinions, we answer three questions in this study. This study was designed to assist future analysts in developing a real model that can easily and accurately forecast consumer evaluation through sentiment analysis. It can also help Netflix understand the emotions of its subscribers. Real-world models can be developed and improved to provide better services and serve consumers.

From this analysis, which is a positive analysis, we discover that the subscribers love the Netflix production. And then the negative tweets to know which things subscribers hate. so that Netflix can improve itself. to provide the desired service to customers. From this analysis, we saw the percentage between positive, negative, and neutral tweets, and it's obvious that the positive tweets are the higher.

To conclude, to determine people's opinions, we built a relatively accurate sentiment analysis. We were able to identify which things subscribers hate and which they love by processing and narrowing the features of the Netflix dataset by identifying the top 10 keywords the audience uses. The most uses keyword about Netflix in twitter for both positive and negative were "Netflix", "watch" and "premium". From this analysis we discover the high frequency between people start from 0.763.

6.1 Recommendation Based on the Findings

- Reduce the cost of membership.

- Use advertisements to compensate for the revenue loss.

- Investigate other areas for development.

- Create a more unique and local substance.

- Increase the security of the account with things like fingerprints or facial recognition. Multi-factor authentication for subscriber authentication

- Deal with more production. studio to increase production

- Increase website security to prevent other websites from stealing the content.

- continues to use sentiment analysis to increase customer satisfaction.

6.2 Limitations

There are a few restrictions that must be recognized even though this study may help the Netflix to create something much better, higher, and stronger. The restrictions of the consideration incorporate:

(a) Data access: we scraped data from Twitter but not from their other social media platforms, such as Instagram, Twitter, WhatsApp, and others. We could scrape data from it.

Data collection: RapidMiner is utilized in this study to conduct the examination portion, but because of the update the software was going through in time of conduction the study we face some problem in running the process on different pc.

(d) Pre-procession: after collecting the data, we had a hard time cleaning up unwanted comments, especially the ones that had emoji in them.

Recommendations for Future Research

1. Collecting data from more sources.

2. Scrape data from various languages to have a higher scope for people's opinions from various languages.

3. Future investigations may also investigate and change the specific strategy for gathering and analyzing data to allow the method to run smoothly.

References

[1] K. L.Sumathy and M. Chidambaram, "Text Mining: Concepts, Applications, Tools and Issues An Overview", *International Journal of Computer Applications*, vol. 80, no. 4, pp. 29-32, 2013.

[2] T. Matsumoto, W. Sunayama, Y. Hatanaka and K. Ogohara, "Data Analysis Support by Combining Data Mining and Text Mining," *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2017, pp. 313-318, doi: 10.1109/IIAI-AAI.2017.165.

[3] P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review", *ICST Transactions on Scalable Information Systems*, vol. 0, no. 0, p. 159623, 2018. Available: 10.4108/eai.13-7-2018.159623.

[4] H. Kwon, H. Ban, J. Jun and H. Kim, "Topic Modeling and Sentiment Analysis of Online Review for Airlines", *Information*, vol. 12, no. 2, p. 78, 2021. Available: 10.3390/info12020078.

[5]D. Hussein, "A survey on sentiment analysis challenges", *Journal of King Saud University - Engineering Sciences*, vol. 30, no. 4, pp. 330-338, 2018. Available: 10.1016/j.jksues.2016.04.002.

[6] "Twitter Announces First Quarter 2022 Results", 2022. [Online]. Available: https://s22.q4cdn.com/826641620/files/doc_financials/2022/q1/Final-Q1%E2%80%9922-earnings-release.pdf. [Accessed: 04- Jul- 2022].

[7]H. Thakkar, D. Patel, '' *Approaches for Sentiment Analysis on Twitter: A State-of-Art study''*, 2015.

[8] "Netflix, Inc. NasdaqGS: NFLX FQ1 2022 Pre Recorded Earnings Call Transcripts", 2022. [Online]. Available: https://s22.q4cdn.com/959853165/files/doc_financials/2022/q1/Netflix,-Inc.,-Q1-2022-Pre-Recorded-Earnings-Call,-Apr-19,-2022.pdf. [Accessed: 04- Jul- 2022].

[9] "About Netflix - Homepage", *About Netflix*, 2022. [Online]. Available: https://about.netflix.com/en. [Accessed: 04- Jul- 2022].

[10] "Amount of original content titles on Netflix 2022 | Statista", *Statista*, 2022. [Online]. Available: https://www.statista.com/statistics/883491/netflix-original-content-titles/. [Accessed: 04- Jul- 2022].

[11] Sasi, S. and F. Alsaqer, A., 2017. Sentiment Analysis of Movie Review using Machine Learning Approach. IJOSTHE, 5(1), p.10.

[12] N A Deraman, A. G. (2022). A Social Media Mining Using Topic Modeling and. IOP Conference Series: Earth and Environmental Science (pp. 1-10). IOP Publishing.