

Table of Contents

1.0 Introduction..... Error! Bookmark not defined.

1.1Data mining	Error! Bookmark not defined.
1.2Text Mining vs. Data Mining:	Error! Bookmark not defined.
1.3 Why is Sentiment Analysis important when Anal	yzing Social Media?Error! Bookmark not defined.
1.4 flowchart for text mining process	Error! Bookmark not defined.
2.0 Tools used in our project	Error! Bookmark not defined.
2.1 Rapidminer tool	Error! Bookmark not defined.
2.2Facepager tool	Error! Bookmark not defined.
3.0 ETL (Extract, Transform, lo	oad)for facebook data Error!
Bookmark not defined.	
4.0 Cleaning Data with excel	Error! Bookmark not defined.
5.0 training data process	Error! Bookmark not defined.
6.0 Load data by Rapidminer	Error! Bookmark not defined.
7.0 ETL (Extract, Transform, lo	oad)for twitter data <u>50</u>
8.0 problems we encountered in defined.	the projectError! Bookmark not

9.0 Solutions and recommendationsError! Bookmark not defined.

Introduction

Data mining is the computational process of discovering patterns in large data sets and establish relationships to solve problems through data analysis involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of data mining is to extract information from a data set and transform it into an understandable structure for further use.

Text Mining vs. Data Mining:

Text mining and data mining are often used interchangeably to describe how information or data is processed.

data mining, which we can define as the discovery of knowledge from structured data (data contained in structured databases or data warehouses.) Today the majority of available business data is unstructured information (articles, website text, blog posts, etc.). The presence of unstructured information makes it more difficult to effectively perform knowledge management activities using traditional business intelligence tools.

The discovery of knowledge sources that contain text or unstructured information is called "text mining". So, the main difference between data mining and text mining is that in text mining data is unstructured Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature level, i.e. whether the expressed opinion in a document, a sentence or an entity feature is positive, negative, or neutral.

The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by precision and recall.

If a program were "right" 100% of the time, humans would still disagree with it about 20% of the time, since they disagree that much about any answer

Our graduation project is about sentiment analysis; our idea is to keep handle on how everyone feels about any brand.

For large companies with thousands of daily mentions and comments on social media, news, sites and blogs, it's extremely difficult to do this manually .to combat this problem, sentimental analysis software are necessary. this software's can be used to evaluate the people's sentiment about any brand or personality.

And we choose **Walmart** as a domain to apply sentimental analysis.

So our problem is organization development, we want to make sentiment analysis in order to keep up with people opinions, know what they want what they said about our services and products in order to help us to improve our companies, know people trend, control our website and social media.

Why is Sentiment Analysis important when Analyzing Social Media?

- Determine marketing strategy
- Improve campaign success
- Improve product messaging
- Improve customer service
- Test business KPIs
- Generate leads

Typical text mining process



Tools that we used in our project:



Rapidminer tool:

is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics.

It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization. RapidMiner is developed on an open core model.



Facepager tool:

- a tool to simplify the process of gathering data from JSON-based APIs without the use of programming languages or predefined scripts, while leaving large degrees-of-freedom to the user. Thus, we do not restrict any API-endpoints and allow "useless" requests.
- a tool to support the step of "data collection" on a low level
- a tool to document the process of data collection, i.e. errors occurring in the process (on both the side of the API and locally, f.e. ill-defined requests).
- a tool that targets researchers/scientific purposes, rather than other audiences like market researchers or other commercial uses.

Find your Facebook ID:

Tool turn our Url in to numerical ID



To find your Facebook personal numeric ID for fb:admins, social plugins, and more, enter your Facebook personal profile URL below:



And this is a simplest flow chart that summarized our work:



ETL (Extract, Transform, load) For Facebook Data

Facepager5.3.

We use the Facepager tool to import the data from Facebook. **Facepager** was made for fetching public available data from Facebook, twitter, YouTube and other JSON-based API. All data is stored in a SQLite database and maybe exported to csv.

18 Facepager 3.6	
🕞 Open Database 📚 New Database 🗍 Export Data 💽 Add Nodes 🗙 Delete Nodes 💈	C Presets D Help
	Add Column Add All Columns
Object ID Object Type Query Status Query Time Query Type	Key Value Data View
Nodes View	Custom Table Columns (one key per line) Column Setup
Facebook Twitter Conorio Files Twitter Streaming Settings	Status Log
Resource Image: Second parameters Node level Parameters Query Setup Maximum pages 1 Parameters Access Token Image: Login to Facebook	1 Inpacked 1 Status View ata C
C:/Users/Jakob/Facepager/Data/test.db	Timer stopped 0 node(s) selecte



Cleaning Data with excel

After we export the data in the excel files we notice that it is noisy as shown below:

FILE HOME INSERT PAGE LAYOUT F	ORMULAS DATA REVIEW	VIEW								Sign in
X Cut Calibri 11 B Copy - B I U - △	A [*] A [*] ≡ ≡ ≡ ≫· ⊨ ↔ • ▲ • ≡ ≡ ≡ € € €	Wrap Text	General \$ - % + *%	Conditional	Format as	Cell Ins	sert Delete Fo	Fill • Clear •	m * AT	d &
Clipboard 72 Font	rs Alignm	ent 5	Number	ъ, с	Styles		Cells		Editing	^
A1 • E × ✓ fx id;"parer	nt id";"level";"objectid";"object	type";"querystatus";"q	uerytime";"queryty	pe";"from.name	e";"messag	e";"likes.su	ummary.tota	l count";"comme	ents.summary.t	otal cou 👻
		10.042 (1.24		M	N	0			с т	
1 Idda second idda Hansalla Habiastid Alabiastication		nutrina Pullfram nama Pul	n L	MI Control or	IN	0	P	4 P	lasa"."un datad	time
1 lo; parent lo; level; objectio; objectivpe;	from our weakly ad http://bit.b	(acbal owners)	nessage ; likes.su	mmary.total_co	unt ; com	nents.sum	mary.total_c	ount ; created_t	ime ; updated	_ume
2 2, 1, 1, get them something theyae in love 3 80°"2"."2" don't adv, it never a you aren't doin	a the cale. Save evenuone the a	rrauation "."".""."2018	-02-07105-10-124000	0, 0 [#] - ^{##}						
4 81*"2"-"2" not "-""-"2018-02-06T15:52:51+000	g me sale. Save everyone me a	gravation., , , 2020	-02-07105.10.121000	~ .						
5 82:"2":"2":"10156306969234236 1015630764059	4236":"data":"fetched (200)":"20	18-02-09 19:07:44 54600	0":"Facebook: <post< td=""><td>>/comments":"</td><td>":"I probab</td><td>ly gonna to</td><td>get that for</td><td>my kids","","","?</td><td>018-02-05T21:2</td><td>8:11+0000":</td></post<>	>/comments":"	":"I probab	ly gonna to	get that for	my kids","","","?	018-02-05T21:2	8:11+0000":
6 83:"2":"2" just no wi start from scratch What	it do I need to have at the registe	r? Do I need an app? Si	hould I just say the h	ell with it and	shop some	one where	else???":"":	":"2018-02-05T19	:33:47+0000":""	
7 84:"2":"2" Walmart!":"":"2018-02-05T18:18:5	9+0000":""							,		
8 85:"2":"2" not cash and was told just a flat no	to the return. They said they did	In't have to take it back	if it was over a certa	ain \$ amount an	d refused t	o tell me w	hat the \$ am	ount was. My ret	turn was not a v	ery high Si
9 86:"2":"2" IL 60620 (was rude am a grown w	oman they need to start treating	people with some resp	ect there.":"":"":"2	018-02-05T23:06	5:14+0000":	**		1		
10 87;"2";"2";"10156306969234236 1015630715896	4236";"data";"fetched (200)";"20	18-02-09 19:07:44.87800	0";"Facebook: <post< td=""><td>>/comments";"</td><td>";"Good tre</td><td>at for Vale</td><td>ntine's Day.</td><td>;"";"";"2018-02-0</td><td>5T17:16:25+000</td><td>0";""</td></post<>	>/comments";"	";"Good tre	at for Vale	ntine's Day.	;"";"";"2018-02-0	5T17:16:25+000	0";""
1 88;"2";"2" but recen but I have let alone the co	ourtesy of a reply or credit to my	account. It would be ni	ce if I could receive	the \$1 earned a	as well as a	response f	rom someon	e";"";"";"2018-02	-07T23:56:16+0	000";""
2 89;"2";"2" 2018 at th I am a per Since she o	did ni she jump. I was able the ca	rt w and I told I do not	re so how cc if I did	I don't rer b	ut becau or	what e ca	lled me an w	ant c Ever hear	To many (welc	ome to Wa
3 90;"2";"2" but I triec in a nutsh that store Tx st	ores even thos and comp are re	ject I had an e despite	tl different and ha	d c because (b	ecause ; th	is is not an	d for employ	ees to be better	trained in the	corporate p
4 91;"2";"2" and our si had to cal we are going to	Aldi's or Sav-a-lot. They can get	our money from now o	n and I actually mea	an it. Whomeve	r came up v	vith the ide	a of those se	an and go needs	fired. Because	this is one
5 92;"2";"2" put in CC info and it said it was sold	out. I don't think they are actual	y selling them at all. I h	ad it in cart. Bumme	d.";"";"";"2018-	02-07T14:3	1:33+0000"	;***			
6 93;"2";"2" and I left and I was even buy a gift	tcard but the w especially I calle	d cuit's made and lex	p and I don't today.	All of this for \$	18";"";"";"2	018-02-06T	15:35:55+000	0";""		
17 94;"2";"2";"10156306969234236_1015630812800	9236";"data";"fetched (200)";"20	18-02-09 19:07:44.87800	0";"Facebook: <post< td=""><td>>/comments";"</td><td>";"Are you</td><td>guys going</td><td>to fix the ca</td><td>reers page? tried</td><td>logging in and</td><td>it keeps say</td></post<>	>/comments";"	";"Are you	guys going	to fix the ca	reers page? tried	logging in and	it keeps say
18 95;"2";"2" who can I they are just giving me th	e run around on a tv that their re	pair co. said is unfixabl	e WHAT CAN I	DO ?????????;;	"";"";"2018-	-02-06T22:2	6:57+0000";"	•		
19 96;"2";"2" we take it today up there and they t	ell us she needs the box it's bee	n 2 months we don't ha	ve the box anymore	. So now she ha	as a broken	vacuum the	at was purch	ased at Walmart	and she's suppo	ose to eat if
20 97;"2";"2" NY Walmi your web we want 1 we w	vent they were but Walmart has	lost some business he	re! Very disappoint	ed in the store	and service	1";"";"";"20	018-02-09T01	:30:52+0000";""		
21 98;"2";"2";"10156306969234236_1015630892991	9236";"data";"fetched (200)";"20	18-02-09 19:07:44.87800	0";"Facebook: <post< td=""><td><pre>>/comments";"</pre></td><td>";"As I cann</td><td>not send a r</td><td>message or p</td><td>ost to page: have</td><td>you seen the p</td><td>hotograph</td></post<>	<pre>>/comments";"</pre>	";"As I cann	not send a r	message or p	ost to page: have	you seen the p	hotograph
22 99;"2";"2";"10156306969234236_1015630924474	4236";"data";"fetched (200)";"20	18-02-09 19:07:44.87800	0";"Facebook: <post< td=""><td>>/comments";"</td><td>";"DON'T s</td><td>hop on wal</td><td>mart.comI</td><td>made a purchase</td><td>and the decrip</td><td>tion said it</td></post<>	>/comments";"	";"DON'T s	hop on wal	mart.comI	made a purchase	and the decrip	tion said it
23 100;"2";"2";"10156306969234236_101563146449	34236";"data";"fetched (200)";"2	018-02-09 19:07:44.8780	00";"Facebook: <po:< td=""><td>st>/comments";</td><td>;"";"How do</td><td>I print off</td><td>my savings o</td><td>atcher now that i</td><td>t has changed?</td><td>";"";"";"201</td></po:<>	st>/comments";	;"";"How do	I print off	my savings o	atcher now that i	t has changed?	";"";"";"201
وهي-الدانا-مخريشة-walmartt-fb	÷	dia al 6. a tanul dala an.					. 8.88.88.89.8000	0.00.07710-00-00	00001.03	
										1.1.1

	And we find that we can process its form into a better one that the Rapid miner tool can deal with, First: We must remain just the comments column	
	🔍 الصفحة الرئيسية 🛛 إدراج تخطيط الصفحة صبغ بيانات مراجعة عرض	y
		1
_		
7	A1 • Jx 81; 2; 2; 10150306969234236_10150309298519236; data; tetched (200); 2018-02-09 19:07:44.546000; Facebook: <pre>comments; ; waimart is the greatest place on a</pre>	Ŷ
<u>/</u>	A A A A A A A A A A A A A A A A A A A	
2	2 82,"2","2","10156306969234236_10156307640594236","data","fetched (200)","2018-02-09 19:07:44.546000","Facebook: <pre>post >/comments","","1 probably going to get that for my kids","","2018-02-05T21:</pre>	
3	3 83;"2","2","10156306969234236_10156307412574236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Anyone out there help me understand the new Savings Catch</post>	ī
4	4 84,"2","2","10156306969234236_10156307264724236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Good luck getting your items if Walmart ships them via Laser</post>	1
2	3 85, 2, 7, 10150300909234230_10150308004384230, data, Tetched (200), 2018-02-09 19:07:44.878000, Facebook <post>/comments, , Gan vvalman stores set up their own return policies separate 3 86:"2"-"2"-"10156306060234236. 10156307847624236"."data"."data"."data"."data"."data"."data</post>	2
7	7 80; 2; 2; 1010000000000000000000000000000)(
3	3 88;"2","2","10156306969234236 10156312935859236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook:state	r
9	9 89;"2";"2";"10156306969234236_10156310372224236";"data";"fetched (200)";"2018-02-09 19:07:44.878000";"Facebook: <post>/comments";"";"1 was an employee for a very short time. I was called into the</post>	1
0	0 90;"2","2","10156306969234236_10156310291279236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","I don't know if this is the correct place to try and get my proble</post>	ε
1	1 91;"2","2","10156306969234236_10156309156894236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Walmart in Wooster Ohio is now a nightmare and I dread goi</post>	1
2	2 92,"2","2","10156306969234236_10156311618/69236","data"; Tetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments",","1 put the Lego anniversary bundle in cart at 8 this morning</post>	n
4	4 94 "2" "2" "10156306969234236 10156308128009236" "data" "fetched (200)" "2018-02-09 19:07:44 878000" "Facebook <post>/comments" "" "Are you guys going to fix the careers page? tried logging in ar</post>	0
5	5 95;"2","2","10156306969234236 10156310190394236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","1 am having a horrible time with the care plan I purchased with</post>	ŕ
6	6 96;"2","2","10156306969234236_10156310099184236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook <post>/comments","","Since there is no way to place a regular comment I guess ill</post>	F .
7	7 97;"2","2","10156306969234236_10156315810514236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Worst experience ever at Watertown</post>	1
8	8 98;"2","2","10156306969234236_10156308929919236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","As I cannot send a message or post to page: have you seen I and a message or post to page: have you seen I and a message or post to page and the dot of the second second</post>	tl
9	a bit, z, z, to subsubsubsubsubsubsubsubsubsubsubsubsubs	;(А
1	1 101;"2","","10156306969234236 10156311492349236","data","fetched (200)","2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> /// Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200)", "2018-02-09 19:07:44.878000"."Facebook: <pre>post</pre> // Work of pink of the second defined (200) "."	i
2	2 102;"2","2","10156306969234236_10156313350669236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Yo Mac after standing around the deli for five minutes why</post>	у
3	3 103;"2";"2";"10156306969234236_10156312724809236";"data";"fetched (200)";"2018-02-09 19:07:44.878000";"Facebook: <post>/comments";"";"Happy Valentine's Day Walmart. Just wanted to let you know</post>	N
4	4 104;"2","2","10156306969234236_10156311641964236","data","fetched (200)","2018-02-09 19:07:44.878000","Facebook: <post>/comments","","Stood at the Deli counter in the Elk grove store two different</post>	t
5	5 106;72;7;72;10156306969234236_101563102912/9236;7data";7tetched (200)";72018-02-09 19:07:44.878000";7Eacebook: <pre>prov // Comments*;77:1010011 know if this is the correct place to try and get my prov // Advance// Comments*;77:1010011 know if this is the correct place to try and get my prov // Advance// Comments*;77:101012912/9246340096649236*;7001101012912/9246340096649236*</pre>)
0	U 101, 2, 2, 10130300303234230_10130310030043230, data, letched (200), 2016-02-03 19.01.44.878000, Facebook spost?/confinents, 1 just (weeted photos to the Walmart and Lay's twitter account	1
		1

Second:

We found the beginning of each comment contain the same size of unusual data so we remove it using the "mid" function in excel for them all at the same time as "MID (text,128, (LEN(text)-128))" and the result is as shown below:

		首	∑ AutoSun	n - Ay	44	
ste BIU·⊞·☆·▲· ■ = = = = = = = = = = = = = = = = = =	Insert Delete	Format	Fill*	Sort &	Find &	
Format Painter Format Painter Format Vainter Format Vainter Format Vainter Format Vainter Styles			Clear*	Filter * S	ielect *	
Capitolitis a solution in agrees	- CER3			Luning		
$32 \bullet : \times \checkmark f_e$						
A		Б	с	D	ε	
but it definitely helps when Im watching my two grandchildrenT				-		T
but our store only has part time positions and the people already there say their hours are being cut due to the self checkout T						
cashiers and zero cheerful yellow vested helpersT						
elyria and Lorain and no little helpers at any of these three Just curiousT						
food lion						
for The website says in store only for and when I add it to my shopping cart online It has the full retail price on it Will the price change online to at price	ST or do I actual	ly have	to go to the	store and g	et it inste	ad o
for some it is but mostly they just want to be with peopleT						
hopefully rd time is a charm if not x your out rule T						
i am on the East Coast EST and it is exactly am here and I went online at about pm Feb nd						
I dont have any problem at all						
I quickly saw that they were the LOL UL SIS dolls and those are NOT what you guys had showing on your app and NOT what I thought I had purchased Nee	dless to say I wa	s VERY d	disappointer	d Not only v	vere they	the
I think My kiddos would of thought he was Santa too Super sweet T						
I would check on your suppliers as i need to deal with you to take large amount of iTunes gift cards weekly so can i contact any representative by email pl	easeT					
if they can do the job more power to the						
IL employer by the name of Melvin						
Im going to give the XFL another tryT	ites the girl tells	me its s	supposed to) be deliver	ed and to	wai
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after minutes and the status of my order it said preparing order. So I to called Wal Mart and after minutes and the status of my order it said preparing order.						
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order So I to called Wal Mart and after mini- items packages items reached the right destination				_		
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after mini- items packages items reached the right destination Lee Hwy						
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after mini- items packages items reached the right destination Lee Hwy manages to put up bins of bulbs were dropped off by the delivery truck which are totally inappropriate for Florida. Ye Gods what are the people response	ble for buying					
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after mini- items packages items reached the right destination. Lee Hwy manages to put up bins of bulbs were dropped off by the delivery truck which are totally inappropriate for Florida. Ye Gods what are the people responses my husband was off work and could still put together. The item did not arrive yesterday. After talking to someone through your chat yesterday afternoon who means a florida.	ible for buying Jorge informed	l me tha	it he was go	ing to cance	I my orde	a ar
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after mini- items packages items reached the right destination. Lee Hwy manages to put up bins of bulbs were dropped off by the delivery truck which are totally inappropriate for Florida. Ye Gods what are the people response my husband was off work and could still put together. The item did not arrive yesterday. After talking to someone through your chat yesterday afternoon not impressedT.	ible for buying Jorge informed	I me tha	it he was go	ing to cance	l my orde	
Im going to give the XFL another tryT It was suppose to be delivered on Saturday But every time I checked the status of my order it said preparing order. So I to called Wal Mart and after mini- items packages items reached the right destination Lee Hwy manages to put up bins of bulbs were dropped off by the delivery truck which are totally inappropriate for Florida. Ye Gods what are the people respons my husband was off work and could still put together. The item did not arrive yesterday. After talking to someone through your chat yesterday afternoon not impressedT pm I visited your Mays Landing NJ store. I was shocked to see the store entirely covered with yellow caution tape closing off most of the store access for	ible for buying Jorge informed shopping It appe	I me tha eared th	it he was go nat Walmart	ing to cance did not war	l my orde t me to s	hop

After that we found that the most of comments had a "T" letter in the end so we decide **Third**:

to remove it also with excel in a quick useful way using the "if" condition and the "mid " function as "if(RIGHT(text)="T", MID (text,1, LEN(text)-1), text)" and the result is as shown below:

ABC Spelling Thesaurus Proofing Proofing Proofing Proofing Smart Language Comments Comm	Workbook :Ranges
A4 🔹 🗄 🔀 A4 I dont have a smart phone so I wont be able to spend the money at a location since the Bluebird	card is going away Yes 🔹 👻
	A
1 but never shipped my box	
2 but never shipped my box	
3 I do not know what the problem is	
4 I dont have a smart phone so I wont be able to spend the money at a location since the Bluebird card is going away Yes	
5 and the item was never shipped	
6 as I sit here crying thinking if only more people were like this in our world and not just celebrities Amazing work Kelly	
7 at the Atlanta Texas store after I was returning from lunch I was told by another employee that I refused an electric cart to a customer	
8 baked goods	
9 but bet you one thing	
10 but it definitely helps when I'm watching my two grandchildren	
11 but our store only has part time positions and the people already there say their hours are being cut due to the self checkout	
12 cashiers and zero cheerful yellow vested helpers	
13 elyria and Lorain and no little helpers at any of these three Just curious	
To the website says in store only for and when Facult to my shopping cart online it has the full retail price on it will the price change	online to at princs) of dollactually have to go to the store and get it instead o
To some its but mostly they but want to be with people	
1 independing to differ a diaminimit for A your out rule	
10 Tain of have any problem at all	
20 Louickly saw that they were the LOLUL SIS dolls and those are NOT what you guys had showing on your app and NOT what I thought I had	ad purchased Needless to say I was VERY disappointed Not only were they the
21 I think My kiddos would of thought he was Santa too Super sweet	
22 I would check on your suppliers as i need to deal with you to take large amount of iTunes gift cards weekly so can i contact any represent	ntative by email please
23 if they can do the job more power to the	Activate Windows
ا الداتا-كاملة-فيسبوك (٠)	Go to Settings to activate Windows.
Ready	III II

Here we notice that there is some of words written in a wrong form, then we search about this problem and find that we can solve it by

using the **spelling process** (Corrects how the word is typed) in excel as the following:

For example, as shown in the picture above the word "don't" is not write in the correct way, also in line '13' the beginning of the word must be in capital letter "elyria" and so on in the remaining data, spelling process can solve these problems,

Then the list shows the wrong word and suggestions to choose the correct one. We click to the correct word "don't" and click change to change it to the correct form as shown:

ABC III 🛈 🛼 📩 🏹 🏹 🖓 Show/Hide Comm	nent 🔲 🧱 Protect and Share Workbook	
Saullies Theorem Smarth Translate New Delate Devices Next Towner	nts Destant Destant Changes	
Lookup Comment	Sheet Workbook Workbook Strack Changes *	
Proofing Insights Language Comments	Changes	^
I don't have a smart phone so I wont be able	to spend the money at a location since the Bluebird card is going away Yes	~
	A	
1 but never shipped my box		
2 but never shipped my box		
3 I do not know what the problem is		
4 I don't have a smart phone so I wont be able to spend the money at a location since	e the Bluebird card is going away Yes	
5 and the item was never shipped		
6 as I sit here crying thinking if only more people were like this in our world and not j	ust celebrities Amazing work Kelly	
7 at the Atlanta Texas store after I was returning from lunch I was told by another em	Microsoft Excel X 1er	
8 baked goods		
9 but bet you one thing	Spell check complete. You're good to go!	
10 but it definitely helps when I'm watching my two grandchildren		
11 but our store only has part time positions and the people already there say their	OK	
12 cashiers and zero cheerful yellow vested helpers		
13 elyria and Lorain and no little helpers at any of these three Just curious		
14 food lion		
15 for The website says in store only for and when I add it to my shopping cart online	It has the full retail price on it Will the price change online to at pm CST or do I	actually have to go to the store and get it instead o
16 for some it is but mostly they just want to be with people		
17 hoperully rd time is a charm if not x your out rule		
10 I dont have any problem at all	20110	
20 Louickly saw that they were the LOLUL SIS dells and these are NOT what you guys h	and chowing on your ann and NOT what I thought I had nurchasod Noodloss to sa	av Lwas VERV disappointed Not only wore they the
21 I think My kiddos would of thought he was Santa too. Super sweet	ad showing on your app and NOT what i thought thad purchased Needless to sa	is it was very disappointed not only were they the
22 I would check on your suppliers as i need to deal with you to take large amount of i	Tunes gift cards weekly so can i contact any representative by email please	
23 if they can do the iob more power to the	rance but on an a configure on the online on the big contained by childin big ase	Activate Windows
	: T7T	Go to Settings to activate Windows.
Reauy		

"spell check complete. You're good to go!"

The same we are used to correct the remaining data.

The data is cleaned, but we want to remove the empty rows using the "filter example" process in rapid miner:



Here the data is cleaned enough to enter the processes of rapid Mainer.

The creation of training data:

In the beginning, after understanding and studying data, we work on training data on the basis that the data relating to the services provided by Walmart such as customer service, technical support, online purchasing, employee handling, without looking at the data that belong to the products sold by Walmart and not produced by himself, so that there is transparency and credibility in the process of analysis of data that belong to this type of work, the data was taken randomly and then categorize them to positive, negative and neutral manually, we tried to some extent for data to be equally in the three classes(positive,negative,neutral), The data was split between the project members and each one of us did a piece of work, study, analyze and classify it. Then we performed a review of each part by the rest of the team members to verify the validity of the classification. As for the number of data selected on the basis of the experiment, we try to experiment more than a number and we apply the model and finally we have stabilized on the number that gives the highest accuracy.

Load data by Rapidminer:

Sample of train data that we use to major the accuracy, in first step we use to class to determine if the comment is positive or negative.

A	15 \mathbf{v} : $\times \checkmark f_{\mathbf{x}}$	
	Α	В
1	text	result
2	Walmart very nice	positive
3	I love shopping at Walmart	positive
4	you came to their aid Thank you for what you do Walmart	positive
5	I always receive friendly and helpful service from the Walmart in Marinette	positive
6	We went and enjoyed the treats at the Perrysburg	positive
7	Walmart is the worst	negative
8	worst customer service ever Walmart online chat wont help me	negative
9	Walmart provides the WORST customer service and with each new change in their return policy it results in them stealing from you	negative
10	Worst experience ever at Walmart	negative
11	Hate Walmart	negative
12	Walmart does not like my comment because it was a bad experience so they marked it as spam	negative

First: we use **Read Excel** operator:

This operator reads an Example Set (sample data) from the specified Excel file, our excel file contain the train data we use, we use import configuration wizard for loading data from excel file to rapid miner Here we see that our excel file is upload on the tool, if we click run we can check that and see our data



Secondly:

Set role: This Operator is used to change the role of one or more Attributes.

Parameters

Attribute name: the name of the Attribute which role should be changed. The name can be selected from the dropdown menu or manual typed. (result as we name in our train file).

Target role: The target role of the selected Attribute is the new role assigned to it.

Label: This is a special role. An Attribute with the label role acts as a target Attribute for learning Operators. The label is also often called 'target variable' or 'class'.



Thirdly:

Nominal to text (To specify which column is a text column, since Rapidminer "Process Documents..." Operators work only on text data.



Fourthly:

We use the Process documents from data operator is used to create word vectors from text attributes, this is a nested operator that contain sub operators inside it.

TF–IDF stands for term frequency–inverse document frequency. It is a numerical statistic which reflects how important a word is to a document in a collection, and it is often used as a weighting factor.



The Tokenize operator tokenizes documents, and we select in the parameters of this operator to tokenize at non letters so that each time a non-letter is found it shall denote a new token, therefore splitting a text into words (This operator splits the text of a document into a sequence of tokens)

The Filter Stopwords (Dictionary) operator applies a stopword list from a file.

Stopwords are words which are filtered out prior to, or after, processing of natural language data (text)

e. For example, some of the most common stopwords for search machines are: the, is, at, which, and on.

) Words that do not matter when parsing text(

The Filter Tokens (by Length) operator, filters tokens based on their length. In its parameters we select the min chars of a token to be 3 (thus

removing single letter words), and the max chars of a token to be 20 which is safe enough to say that words consisting of 20 chars are probably gibberish.

Stemming also known as lemmatization is a technique for the reduction of words into their stems, base or root. Many words in the English language can be reduced to their base form or stem e.g. like, liking, likely, unlike belong to like.

The Transform Cases operator transforms the cases of all characters. In its parameters we choose to transform all characters to lower case.





Number of train data sets= 500 row (comment, class)

K (number of folds) =10, number of rows (data set size) =500

500/10=50

50	50	50	50	50
50	50	50	50	50

The Cross Validation Operator is a nested Operator. It has two sub processes: A Training sub process and a Testing sub process. The Training sub process is used for training a model. The trained model is then applied in the Testing sub process. The performance of the model is measured during the Testing phase.

In the beginning, cross validation is a process that divides train data into parts, K is chosen as 10, and according to many experiments for more numbers we discover that number 10 is the one that gives the highest accuracy. In addition, we investigated several files of the subject and the majority advice to choose the number 10 to give better results and better learning ability. **cross validation works as follows:**

Divide the 10 groups into two parts as shown in the table above. The first section contains (9/10) groups which contains the part that the model will learn from, and the second section contains one set (1/10) which is the part that we will do a test on it and show us a certain accuracy, after that this process will make reverses groups that take 9 different groups for learning and one group for testing process and returns to calculate the accuracy in the same way and so on (the process does the switching process until it finishes all the groups (train, test) and takes the average of All the accuracy that appeared in all operations, and shown to us as a final result rate and this is the result on which to calculate whether the accuracy is high or not.

And this process contains sub processes as shown below:



Number 1 is the (training part) that contains the Classification model (decision tree).

Number 2 is the (testing part) that contains: -

1- The apply model (This Operator applies a model on an Example Set (train set).

2- The performance operator (This operator is used for performance evaluation. It delivers a list of performance criteria values. These performance criteria are automatically determined in order to fit the learning task type.)

After We click on RUN (\triangleright) we can notice the program run by see the cross validation operator

accuracy: 83.52% +/- 4.6	ccuracy: 83.52% +/- 4.86% (mikro: 83.52%)					
	true positive	true nigative	class precision			
pred. positive	204	23	89.87%			
pred. nigative	66	247	78.91%			
class recall	75.56%	91.48%				

Through analyzing comments (positive, negative), we discovered a comments that is not positive and not negative(neutral), so we decide to add a new class called: **neutral**.

Here is a sample of train data that contains three classes (positive, negative, neutral) that we use it to measure the accuracy, to ensure that the machine learning is applied on training data (pos, neg, nut) in the correct way or not (does the model learn right from the human or not?).

rext	result
so well organized Thankful for all of the great employees	positive
which is great for me because I have a baby Thanks Walmart You are doing a great job	positive
you came to their aid Thank you for what you do Walmart	positive
love shopping at WalMart even when crowded especially toys and books and this picture is awesome	positive
love the Walmart where I live They have everything	positive
Nalmart is the greatest place on earth	positive
NalMart Really not cool	nigative
Nalmart is the worst	nigative
WORST EXPERIENCE EVER THANKS WALMART from a former customer	nigative
After waiting an eternity for my online purchase and bad treatment for their customer service	nigative
was My local walmart had a very very bad experience on January	nigative
but Walmart has lost some business here Very disappointed in the store and service	nigative
Walmart online is useless	nigative
shop at Walmart	nutral
Nalmart sets the standard In this case	nutral
Nalmart im trying to buy a gift set that is in your black friday ad but it shows a different price Can you help me please Pease	nutral
Do Walmart have whatsapp group chat	nutral
Hello I was wondering if Walmart does international shipping to Argentina	nutral
Nalmart I Was wondering if you guys carry security jackets?	nutral

After that we build the model (with different operators) that described in details in the previous pages (in training data for 2 classes), and **run** the process to show the results below:

2	-	€		Vi	ews: Design	Results				
	Result Hist	ory ×			🎇 Perforn	nanceVector (Per	formance)	х		
	📒 ExampleSet (Cross Validation) 🛛 🛛 🛛			📒 ExampleSet (Cross Validation) 🛛 🛛 🛛				🏆 Tree (Decision Tree) 🛛 🗙		
	ExampleSet (809 examples, 5 special attributes, 1777			regular attributes)			Filter (809 / 809 examples):		all	
Data	Row No.	positive	prediction(p	confidence(nutral)	confidence(positive)	confidence(aai	aand	abl	abse
	1	positive	nutral	0.462	0.077	0.460	0	0	0	0
	2	positive	positive	0.085	0.872	0.043	0	0	0	0
Σ	3	positive	nutral	0.462	0.077	0.460	0	0	0	0
Statistics	4	positive	positive	0	1	0	0	0	0	0
	5	positive	positive	0	1	0	0	0	0	0
	6	positive	positive	0	1	0	0	0	0	0
Charts	7	positive	positive	0.085	0.872	0.043	0	0	0	0
	8	positive	nutral	0.462	0.077	0.460	0	0	0	0
	9	positive	positive	0	1	0	0	0	0	0
Advanced	10	positive	positive	0.085	0.872	0.043	0	0	0	0
Charts	11	positive	positive	0	1	0	0	0	0	0
	12	positive	nutral	0.462	0.077	0.460	0	0	0	0
	13	positive	positive	0	1	0	0	0	0	0
Annotations	14	positive	positive	0	0.986	0.014	0	0	0	0
	15	positive	positive	0	0.986	0.014	0	0	o Acti	vate

Here the picture shows the **positive column** (training data) and **prediction column)** the result predicted by the model), and the percent of confidence for each class. For example, in the first row the training data is (positive) and the prediction is (neutral), this means that the model is not learning well from the training data) the result that the model predict is wrong).



accuracy: 56.99% +/- 3	3.33% (mikro: 56.98%)			
	true positive	true nigative	true nutral	class precision
pred. positive	198	17	7	89.19%
pred. nigative	0	0	0	0.00%
pred. nutral	71	253	263	44.80%
class recall	73.61%	0.00%	97.41%	

The accuracy from training data in 3 classes (positive, negative, neutral) is 56.99 which is not good enough, so we want to use a new way that increase the accuracy and that make the model distinguish between classes and learn well from training data. (the details about this in the next pages).

As we see above the accuracy of the learning is not good enough so we found that we should increase it by **First**:

Test the model with two classes (Neutral, Unnatural) and measure the accuracy.

Second:

Test the model with another two classes (positive, negative) and measure the accuracy and then compute the average between two measure accuracy as we will see below:

1	text	result
2	Walmart is the worst	unnutral
3	WalMart Really not cool	unnutral
4	I hate your page I wanted the phone number for my walmart couldnt get it the store finder has no button to push when you put your city and state	unnutral
5	The staff at the new Walmart in my area arent helpful at all They displayed how much they hate working there all over their faces Please correct that before too late	unnutral
6	have a very bad experience at Fairfax Walmart Supercenter	unnutral
7	Andrea did you look in the outdoor section? At my Walmart they had the Holiday CD display by the blow up Christmas stuff in the Home and Garden Section	nutral
8	Walmart in roxboro NC will take them	nutral
9	Walmart dont know when theyll get them in again	nutral
10	hello walmart how can i change my shipping address	nutral
11	When I click on refill prescription in My Walmart pharmacy account	nutral

Using **First** case with (Neutral, Unneutral)-Training Data:



We use this model –that is explained previously- to compute the accuracy of the previous training data, and we find notice that the accuracy is 76.75 as we see here:

accuracy: 76.75% +/- 4.0	60% (mikro: 76.72%)		
	true unnutral	true nutral	class precision
pred. unnutral	157	15	91.28%
pred. nutral	110	255	69.86%
class recall	58.80%	94.44%	

Here we will execute the classification as (neutral and unneutral) to whole Walmart Facebook data so:

First: we use the read excel process to load all Walmart Facebook data in rapidminer as shown below:



Second:

we use the Nominal to text process again as we made in the training data (To specify which column is a text column, since Rapidminer "Process Documents..." Operators -that will be used in the next process - work only on text data.

As we will see below:



Third:

we used the (Process documents from data) process that we explain it previously to clean the text using a lot of sub processes inside of it as we will see below:



And this process as previously mentioned it contain a lot of sub processes as we see below:



The Tokenize operator tokenizes documents, and we select in the parameters of this operator to tokenize at non letters so that each time a non-letter is found it shall denote a new token, therefore splitting a text into words (This operator splits the text of a document into a sequence of tokens)

The Filter Stopwords(Dictionary) operator applies a stopword list from a file.Stopwords are words which are filtered out prior to, or after, processing of natural language data (text)

e. For example, some of the most common stopwords for search machines are: the, is, at, which, and on.

) Words that do not matter when parsing text(

The Filter Tokens (by Length) operator, filters tokens based on their length. In its parameters we select the min chars of a token to be 3 (thus removing single letter words), and the max chars of a token to be 20 which is safe enough to say that words consisting of 20 chars are probably gibberish.

Stemming also known as lemmatization is a technique for the reduction of words into their stems, base or root. Many words in the English language can be reduced to their base form or stem e.g. like, liking, likely, unlike belong to like.

The Transform Cases operator transforms the cases of all characters. In its parameters we choose to transform all characters to lower case

Forth:

we execute the model using the (Apply Model) process, This Operator applies a model on an Example Set. As shown below:


And then we must take all the predicted comments that done by the rapidminer using the Write excel process as follows:



The result we see is as follows after we click run:





PerformanceVector

PerformanceVector:
accuracy: 76.75% +/- 4.60% (mikro: 76.72%)
ConfusionMatrix: True: unnutral
nutral unnutral: 157–15
nutral: 110 255
precision: 70.21% +/- 4.61% (mikro: 69.86%) (positive class: nutral)
ConfusionMatrix: True: unnutral nutral unnutral: 157 15
nutral: 110 255
recall: 94.44% +/- 3.80% (mikro: 94.44%) (positive class: nutral)
ConfusionMatrix: True: unnutral nutral unnutral: 157
15
nutral: 110 255
AUC (optimistic): 0.978 +/- 0.015 (mikro: 0.978) (positive class: nutral)
AUC: 0.767 +/- 0.045 (mikro: 0.767) (positive class: nutral)
AUC (pessimistic): 0.555 +/- 0.087 (mikro: 0.555) (positive class: nutral)

	-	5			Views:	Design	Results					🗿 Need help? 🔻
% Perfo	ormanceVector (P	erformance)	×	ExampleSet (0	Cross Validatio	n) ×	Ex:	ampleSet (Cross Valio	lation)	×	Repository ×	
Result	History ×		Exam	pleSet (Apply M	odel (2))	×		Tree (Decision Tr	ee) ×		G Add Data	= -
-	ExampleSet (5-	450 examples, 3 spe	ecial attributes, 140	8 regular attributes	;)		Filter (5,	450 / 5,450 examples):	all	•	Samples	
Data	Row No.	prediction(r	confidence(confidence(aai	aand	abl	absentmind	absolut	acci	▶ 📕 DB	
Data	1	nutral	0.239	0.761	0	0	0	0	0	0 ^	The grad pro (hp)	
	2	nutral	0.239	0.761	0	0	0	0	0	0	idsi (hp - v1, 4/7/18 11:0)	ے فینیز کے AM - 10 kB
Σ	3	nutral	0.239	0.761	0	0	0	0	0	0	tajroba (hp - v1, 3/2	5/18 8:25 PM - 13 kB
Statistics	4	nutral	0.239	0.761	0	0	0.425	0	0	0	🛷 testtttt (hp - v1, 3/1/	18 10:41 AM - 18 kB)
	5	nutral	0.239	0.761	0	0	0	0	0	0	🗳 yas (hp - v1, 2/11/18	9:41 AM - 1 kB)
	6	nutral	0.239	0.761	0	0	0	0	0	0	yasmeen (hp - v1, 1)	2/17/18 7:15 PM - 9 k
Charts	7	unnutral	1	0	0	0	0	0	0	0	(hp - v1, 2/6/18 7:22	برل فيبرك (PM - 5 kB) مرال فيبرك
	8	unnutral	1	0	0	0	0	0	0	0	🗳 (hp - v1, 2/16/18 10:1	برال کلین (BM - 4 kB) ا
	9	nutral	0.239	0.761	0	0	0	0	0	0	(hp - v1, 3/9/18 7:55	رر ^ی (PM - 6 kB
Advanced	10	nutral	0.239	0.761	0	0	0	0	0	0	(hp - v1, 2/16/18 5:3)	5 PM - 4 kB)
Charts	11	nutral	0.239	0.761	0	0	0	0	0	0	• mytra (hp)	(m - 10 kB)
	12	nutral	0.239	0.761	0	0	0	0	0	0	🕨 🚞 data (hp)	
	13	nutral	0.239	0.761	0	0	0	0	0	0	process (hp)	
Annotations	14	nutral	0.239	0.761	0	0	0	0	0	0	ayl (hp - v1, 12/28/1)	7 12:57 AM - 1 kB)
	15	unnutral	1	0	0	0	0	0	0	0	gra (hp - v1, 12/22/1	7 7:43 PM - 845 byte
	<									>	<	~

		Name	🕨 - Туре	Missing	Statistics Filter (1,4	11 / 1,411 attributes): Search	for Attributes
Data	~	Prediction prediction(result)	Polynominal	0	^{Least} unnutral (700)	^{Most} nutral (4750)	Values nutral (4750),
Σ	~	Confidence_unnutral confidence(unnutral)	Real	0	Min 0.239	Max 1	Average 0.337
lausucs	~	Confidence_nutral confidence(nutral)	Real	0	Min O	Max 0.761	Average 0.663
Charts	~	aai	Real	0	Min O	Max 0.399	Average 0.000
2	~	aand	Real	0	Min O	Max 0.398	Average 0.000
dvanced Charts	~	abl	Real	0	Min O	Max 0.883	Average 0.003
notations	~	absentmind	Real	0	Min O	^{Мах} 0.434	Average 0.000



And secondly after we extract the neutral and unneutral data we take all unnutral data and test it using this training data (positive, negative)training data:

A	15 $\overline{}$: $\times \checkmark f_{x}$			
	А	в		
1	text	result		
2	Walmart very nice	positive		
3	I love shopping at Walmart	positive		
4	you came to their aid Thank you for what you do Walmart	positive		
5	I always receive friendly and helpful service from the Walmart in Marinette	positive		
6	We went and enjoyed the treats at the Perrysburg	positive		
7	Walmart is the worst	negative		
8	worst customer service ever Walmart online chat wont help me	negative		
9	Walmart provides the WORST customer service and with each new change in their return policy it results in them stealing from you	negative		
10	Worst experience ever at Walmart	negative		
11	Hate Walmart	negative		
12	Walmart does not like my comment because it was a bad experience so they marked it as spam	negative		

and using the previous model we find the accuracy 83.31 as shown below

accuracy: 83.31% +/- 4.6	57% (mikro: 83.30%)		
	true positive	true nigative	class precision
pred. positive	201	22	90.13%
pred. nigative	68	248	78.48%
class recall	74.72%	91.85%	

Here we found that the accuracy of using these two processes is (76.62+83.31)/2

=79.9

This result is better than using (positive, negative, neutral) as a training data

And we also we try it using deferent classification models like (SVM, naïve bays) as follows:



Let's start to building our model using the training data that we reached in previous page and make the prediction to the whole Walmart Facebook data.

First: we use the read excel process to load all Walmart Facebook data in rapidminer as shown below:



Second: we use the Nominal to text process again as we made in the training data (To specify which column is a text column, since Rapidminer "Process Documents..." Operators -that will be used in the next process - work only on text data.

As we will see below:



And this process as previously mentioned it contain a lot of sub processes as we see below:

exa D

wor D

wor

exa

4

out

exa 📩

exa

ori 🗋



The Tokenize operator tokenizes documents, and we select in the parameters of this operator to tokenize at non letters so that each time a non-letter is found it shall denote a new token, therefore splitting a text into words (This operator splits the text of a document into a sequence of tokens)

The Filter Stopwords(Dictionary) operator applies a stopword list from a file. Stopwords are words which are filtered out prior to, or after, processing of natural language data (text)

e. For example, some of the most common stopwords for search machines are: the, is, at, which, and on.

) Words that do not matter when parsing text(

The Filter Tokens (by Length) operator, filters tokens based on their length. In its parameters we select the min chars of a token to be 3 (thus removing single letter words), and the max chars of a token to be 20 which is safe enough to say that words consisting of 20 chars are probably gibberish.

Stemming also known as lemmatization is a technique for the reduction of words into their stems, base or root. Many words in the English language can be reduced to their base form or stem e.g. like, liking, likely, unlike belong to like.

The Transform Cases operator transforms the cases of all characters. In its parameters we choose to transform all characters to lower case.

Forth: we execute the model using the (Apply Model) process, This Operator applies a model on an Example Set. As shown below:



Then we click **Run** to see and analyze the result as shown below:









Decision tree





For the Twitter Data ETL (Extract, Transform, load) for twitter Data

we will start with the extraction of Wal-Mart data from twitter, the tool that we use for this purpose_is Rapidminer tool

RapidMiner tool.

Rapid miner is a software platform for data science teams that unites data preparation, machine learning, and predictive model deployment.

We Create a new process in RapidMiner Studio then Here we drag the "search twitter operator " and then enter to the twitter account through the operator and then we choose the keywords that we want to search about like (Walmart , #Walmart, @Walmart , Love Walmart , #love Walmart , hate Walmart , #hate Walmart).

As shown below:

i i		
Search Twitter		ា

Before we run the process, we put it in new excel file through the process "write excel" and put all data we generate on it.



Start ----- > new connection (open Url "Request access token - login
to twitter ", copy access token, test, save all changes) ----- > query ----- > limit ----- > language ----- > write excel ----- > Run

The data we generate is as sown below:

Fil	e Home Ir	isert Pag	e Layout Formulas	Data	Review	View	Q Tell	me what y	ou want te	o do								Sign in	A Sha	are
Past	Cut Copy - Format Painter	Calibri BIL				- 14	📴 Wrap	Text e & Center	Gen	• % •	•.8 •.8	Conditiona Formatting	I Format	as Cell Styles	Insert Delete Format	∑ Aut ↓ Fill	toSum • A • Z ar • Filt	♥ ♀ t& Find & er * Select *		
	Clipboard	G.	Font G			Alignment			1%	Number	G.		Styles		Cells		Editing			^
01		×Z	fx																	~
-			·																	
A	A	В	С	D		E		F	G	н	1	J		К	L	м	N	0	Р	1
1 (reated-At	From-Use	From-User-Id	To-User	To-Use	r-Id		Language	Source	Text	Geo-	Local Geo-L	ocal Retu	veet-()	1					-
2	2018-03-17 00:28:0	4 PETA	9890492.0				-1.0	en	<a href="</td"><td>"I Thank y</td><td>ou @Wa</td><td>almart for re</td><td>emo</td><td>159.0</td><td>974774327840264000.0</td><td></td><td></td><td></td><td></td><td>-</td>	"I Thank y	ou @Wa	almart for re	emo	159.0	974774327840264000.0					-
3	2018-03-17 19:30:0	3 Friends of	19539716.0				-1.0	en	<a href="</td"><td>"r.@Waln</td><td>nart war</td><td>nts to buy a</td><td>robc</td><td>45.0</td><td>975061717020479000.0</td><td></td><td></td><td></td><td></td><td></td>	"r.@Waln	nart war	nts to buy a	robc	45.0	975061717020479000.0					
4	2018-03-13 21:02:3	7 Funko	1378000488.0				-1.0	en	<a href="</td"><td>"I RT & fo</td><td>llow @O</td><td>riginalFunk</td><td>to fo 15</td><td>781.0</td><td>973635458134626000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT & fo	llow @O	riginalFunk	to fo 15	781.0	973635458134626000.0					
5	2018-03-18 19:53:4	7 Agape Lov	3581952077.0				-1.0	en	<a href="</td"><td>"I RT @Ca</td><td>ndace_P</td><td>Parker: My I</td><td>ittle</td><td>107.0</td><td>975430076123804000.0</td><td></td><td></td><td></td><td></td><td>-12</td>	"I RT @Ca	ndace_P	Parker: My I	ittle	107.0	975430076123804000.0					-12
6	2018-03-18 19:53:4	5 Kyle Mato	727284014332150000.0				-1.0	en	<a href="</td"><td>"h RT @Or</td><td>iginalFu</td><td>nko: RT & fe</td><td>ollov 5</td><td>482.0</td><td>975430069656195000.0</td><td></td><td></td><td></td><td></td><td>-</td>	"h RT @Or	iginalFu	nko: RT & fe	ollov 5	482.0	975430069656195000.0					-
7	2018-03-18 19:53:2	6 CB Blog&F	457243990.0				-1.0	en	<a href="</td"><td>"I got a f</td><td>ree Fibe</td><td>r One Supr</td><td>eme</td><td>.0</td><td>975429988819317000.0</td><td></td><td></td><td></td><td></td><td></td>	"I got a f	ree Fibe	r One Supr	eme	.0	975429988819317000.0					
8	2018-03-18 19:53:2	2 presley	2704993662.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429970204987000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429970204987000.0					
9	2018-03-18 19:52:5	4 #EdgarHO	126136711.0				-1.0	en	<a href="</td"><td>"h RT @pin</td><td>nkyjone</td><td>z: I wonder</td><td>if th</td><td>1.0 9</td><td>975429852680487000.0</td><td></td><td></td><td></td><td></td><td></td>	"h RT @pin	nkyjone	z: I wonder	if th	1.0 9	975429852680487000.0					
10	2018-03-18 19:52:2	6 yung meri	229986822.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429735982486000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429735982486000.0					
11	2018-03-18 19:51:2	7 C.	1141492476.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0 9</td><td>975429490498253000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0 9	975429490498253000.0					
12	2018-03-18 19:51:2	5 Litecoin K	956760303257055000.0				-1.0	en	<a href="</td"><td>"h RT @th</td><td>elitesch</td><td>ool: Hey @\</td><td>Waln</td><td>26.0</td><td>975429482541715000.0</td><td></td><td></td><td></td><td></td><td></td>	"h RT @th	elitesch	ool: Hey @\	Waln	26.0	975429482541715000.0					
13	2018-03-18 19:51:0	5 Janay Mic	499448061.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429396780724000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429396780724000.0					
14	2018-03-18 19:50:5	0 ??p	168317247.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429334377824000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429334377824000.0					
15	2018-03-18 19:50:1	7 Fabiansito	806733619406389000.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429196414566000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429196414566000.0					
16	2018-03-18 19:50:1	0 Melancho	349980463.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429167885013000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429167885013000.0					
17	2018-03-18 19:49:5	6 Persian Cu	324091138.0				-1.0	en	<a href="</td"><td>"I RT @ut</td><td>arlingtor</td><td>n: UTA's ow</td><td>n @ı</td><td>16.0</td><td>975429105918403000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @ut	arlingtor	n: UTA's ow	n @ı	16.0	975429105918403000.0					
18	2018-03-18 19:49:5	6 J WOKEST	967819864751587000.0				-1.0	en	<a href="</td"><td>"h RT @Pr</td><td>ogDadT</td><td>/: Dear @W</td><td>alm.</td><td>18.0</td><td>975429105918255000.0</td><td></td><td></td><td></td><td></td><td></td>	"h RT @Pr	ogDadT	/: Dear @W	alm.	18.0	975429105918255000.0					
19	2018-03-18 19:49:3	5 draizy ??	3305778788.0				-1.0	en	<a href="</td"><td>"I RT @TE</td><td>VO_SPR</td><td>ITEEEE: Excu</td><td>use r</td><td>901.0</td><td>975429018768916000.0</td><td></td><td></td><td></td><td></td><td></td>	"I RT @TE	VO_SPR	ITEEEE: Excu	use r	901.0	975429018768916000.0					
20	2018-03-18 19:47:5	4 Shawn	23722811.0	Walmart		1713	7891.0	en	<a href="</td"><td>"I @Walm</td><td>nart @To</td><td>psMarkets</td><td>tops</td><td>.0</td><td>975428597770084000.0</td><td></td><td></td><td></td><td></td><td></td>	"I @Walm	nart @To	psMarkets	tops	.0	975428597770084000.0					
21	2018-03-18 19:47:3	2 Retired N	815447261224337000.0	CDC_DJ_	A	222144	5064.0	en	<a href="</td"><td>"I @CDC_</td><td>DJ_Acco</td><td>unt @Waln</td><td>nart</td><td>.0 .</td><td>975428501825343000.0</td><td></td><td></td><td></td><td></td><td></td>	"I @CDC_	DJ_Acco	unt @Waln	nart	.0 .	975428501825343000.0					
22	2018-03-18 19:47:0	2 Laura Bow	26152056.0				-1.0	en	<a href="</td"><td>"I Thank y</td><td>rou, @tif</td><td>fanymstev</td><td>ens,</td><td>.0</td><td>975428377229386000.0</td><td></td><td></td><td></td><td></td><td></td>	"I Thank y	rou, @tif	fanymstev	ens,	.0	975428377229386000.0					
23	2018-03-18 19:46:5	9 DES	336278605.0				-1.0	en	<a href="</td"><td>"I Okay bu</td><td>ut for rea</td><td>al why is Po</td><td>cahc</td><td>.0 .</td><td>975428363895607000.0</td><td>ivate \</td><td>Windows</td><td></td><td></td><td></td>	"I Okay bu	ut for rea	al why is Po	cahc	.0 .	975428363895607000.0	ivate \	Windows			
24	2018-03-18 19:46:5	5 biorn130	97455889.0				-1.0	en	<a href="</td"><td>"I RT @Or</td><td>iginalFu</td><td>nko: RT & fe</td><td>ollov 5</td><td>482.0</td><td>975428349945418000.0</td><td>n Sottin</td><td>as to activa</td><td>to Windo</td><td></td><td></td>	"I RT @Or	iginalFu	nko: RT & fe	ollov 5	482.0	975428349945418000.0	n Sottin	as to activa	to Windo		
4	Rapid	Miner Data	•									4				- DOLLIN) [Þ.

Then we delete all un useful columns and remain just the tweets.

🖬 🕏 • 🗟 - 후 Book1 - Excel (Product Activation Failed)			a –	. 8	×
File Home Insert Page Layout Formulas Data Review View 🖓 Tell me what you want to do			S	ign in 🙎	. Share
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Delete Format	∑ AutoSur ↓ Fill * ℓ Clear *	n • Av Sort & Filter •	Find & Select *	~
$F7 = : \times \sqrt{f_x}$					
A	В	с	D	E	F
2 Thank you @Walmart for removing all ejiao products from your shelves! Donkeys are bludgeoned as young as 5 months o https://t.co/I3TpNXIIX3					
3 .@Walmart wants to buy a robot bee to produce the food it sells. They should stop selling food grown with bee-kill https://t.co/MuoF3tyj3Q					
4 RT & follow @OriginalFunko for the chance to win a @Walmart exclusive Thanos Pop! https://t.co/jmAfpRA9Oz					
5 RT @Candace_Parker: My little Black Panther! <3 This movie gave Lailaa a new perspective on what superheroes can look like. And @Walmart ga					
6 RT @OriginalFunko: RT & follow @OriginalFunko for a chance to WIN a @Walmart exclusive jade Sixer Pop! #ReadyPlayerOne https://t.co/ljNfo8					
7 I got a free Fiber One Supreme Brownie sample from the #Freeosk at @Walmart https://t.co/lax5RdDkxy					
8 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
9 RT @pinkyjonez: I wonder if there is anything that @peta @ASPCA or the @HumaneSociety can do about the conditions in which @Walmart keeps					
10 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
11 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
12 RT @theliteschool: Hey @Walmart. Just an idea. If you need help, let me know. ?? #litecoin \$ltc https://t.co/mjuQXaB65I					
13 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
14 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
15 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
16 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
17 RT @utarlington: UTA's own @riseacappella helped Cheerios bring a few moments of good to some surprised @Walmart shoppers! https://t.co/Fm1					
18 RT @ProgDadTV: Dear @Walmart my child had to see this in your store yesterday. Please explain to me how it is OK to promote blackface in yo					
19 RT @TEVO_SPRITEEEE: Excuse me earlier at your store I asked your employee what aisle is the hot dogs in and she told me "7 but the buns rig					
20 @Walmart @TopsMarkets tops has better prices on my nutrional stuff i need for the gym then you do! And they give me https://t.co/MVkCiaFIDv					
21 @CDC_DJ_Account @Walmart They never have enough cashiers, the store aisles are packed and it's hard to move around https://t.co/GVIt5t6EN2					
22 Thank you, @tiffanymstevens, for this excellent reporting! We definitely need more point-of-sale medication drop bo https://t.co/XiZDAEUE0n			15		
23 Okay but for real why is Pocahontas white? @Walmart @Disney https://t.co/2RqUNjNKCq	Activ	rate Wind	dows		*
Sheet1 ⊕ : ◀	Go to	Settings to	activate V	Vindows.	Þ
Select destination and press ENTER or choose Paste	Ħ	I II		+	F 100%

here the data that we want to work on it "the tweets" from twitter

We consider to create the training data that will be used to learn the machine and make it predict the positivity and negativity off whole Walmart twitter data so **First** The data is as shown below:

4	A	В
1	Messege	Result
2	Can't understand why someone would leave such a high level position to work at Walmart? One can effect rea	negative
3	#Walmart reminding why I hate Walmart#shit store	negative
4	@_kendrapaige_I hate Walmart in general, but that one is my least favorite.	negative
5	@AFootballLife LoL see this is another example of why I hate Walmart	negative
6	Don't forget corporate America!!! People loveeee to talk about how much they hate Walmart but won't su	negative
7	I love going to Walmart ????	positive
8	I went to Walmart and there was a mom and her lil son and they were "competing" with "I love you!!" "I love YOU" "	positive
9	Thanks for the excellent service @walmart #grocery pickup in Estero. Kevin truly makes you feel like the most important customer of the day.	positive
10	Was working at a pretty nice walmart today and a cute little mouse ran across the floor in front of me. Gave 5 CSMs a heart attack.	positive
11	you are the best walmart go aHEAD	positive
12	morining all	neutral
13	Today I went to Walmart and there was a foundation bottle that was completely empty on the shelf	neutral
14	how are you	neutral
15	Lil Xan is at Walmart and I can't even go see him smh??	neutral
16	RT @MrGoodBeard: Walmart has self checkout you can get your food for free lol	neutral

we take three classes from the beginning because we notice that there are some comments is neutral as shown previously.





Third:

Set role: This Operator is used to change the role of one or more Attributes

Parameters:

Attribute name ->the name of the Attribute which role should be changed. The name can be selected from the dropdown menu or manual typed. (result as we name in our train file).

Target role->The target role of the selected Attribute is the new role assigned to it. Following target roles are possible:

Label: This is a special role. An Attribute with the label role acts as a target Attribute for learning Operators. The label is also often called 'target variable' or 'class'

Dinp		
Read Excel	Set Role	
	ori	

Fourth: we use the Nominal to text process (To specify which column is a text column, since Rapidminer "Process Documents..." Operators work only on text data.

		Views:	Design	Results		
Repository X	Process ×					
G Add Data ≡ 💌	Process				100% 🔎 🔎 🔎	🛛 🗣 🗲 📮
Samples	Process					
📮 DB	Dinp					res
mytra (hp)	Read Excel	Set Role	ominal to Text		5	res
NewLocalRepository (hp)	🖌 fil 📩 out 🕅 🛶 🕻 ex	a 🚺 exa	exa 🎓 exa			
مدروع حوم مرما مير Cloud Repository (disconnected)						
)perators ×						
nominal to tex 🗙						
Blending (2)						
 Attributes (2) Times (2) 						
Nominal to Text						
🍅 Text to Nominal						

Fifth:

We use Process Documents from Data process:



The Process Documents from Data operator is used to create word vectors from text attributes, this is a nested operator that contain sub operators inside it

TF–IDF stands for term frequency–inverse document frequency. It is a numerical statistic which reflects how important a word is to a document

in a collection, and it is often used as a weighting factor. And it contains sub processes like:



The Tokenize operator tokenizes documents, and we select in the parameters of this operator to tokenize at non letters so that each time a non-letter is found it shall denote a new token, therefore splitting a text into words (This operator splits the text of a document into a sequence of tokens)

The Filter Stopwords (Dictionary) operator applies a stopword list from a file.

Stopwords are words which are filtered out prior to, or after, processing of natural language data (text).

e. For example, some of the most common stopwords for search machines are: the, is, at, which, and on.

) Words that do not matter when parsing text(

The Filter Tokens (by Length) operator, filters tokens based on their length. In its parameters we select the min chars of a token to be 3 (thus removing single letter words), and the max chars of a token to be 20 which is safe enough to say that words consisting of 20 chars are probably gibberish.

Stemming also known as lemmatization is a technique for the reduction of words into their stems, base or root. Many words in the English language can be reduced to their base form or stem e.g. like, liking, likely, unlike belong to like.

The Transform Cases operator transforms the cases of all characters. In its parameters we choose to transform all characters to lower case



Sixth: we use the cross validation process as shown below:

The Cross Validation Operator: is a nested Operator. It has two sub processes: A Training sub process and a Testing sub process. The Training sub process is used for training a model. The trained model is then applied in the Testing sub process. The performance of the model is measured during the Testing phase

K=10, we choose k=10 because we discover that 10 is give us a better machine Learning and a better accuracy, the model learn better What is machine learning?

Our project based on machine learning, we choose a train data and classification model and the apply of that classification model on our train data call machine learning Generally, there are 3 types of learning algorithm in our project we use Supervised Machine Learning Algorithms to make predictions we use this machine learning algorithm. Further, this algorithm searches for patterns within the value labels. That was assigned to data points. (we already know the output of our analyses (positive, negative, neutral)).

I alli ua	$\begin{vmatrix} a - b \\ \end{vmatrix}$	010w t	llat ale	Classify	manua	illy			
Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
/		1		1	/		I	I	
	Max ->	• best val	lidation				Max-2	>best lea	arning re

Train data = 500 row that are classify manually

Number of train data sets= 500 row (comment, class)

K (number of folds) =10, number of rows (data set size) =500

500/10= 50

50	50	50	50	50
50	50	50	50	50

In the beginning, cross validation is a process that divides train data into parts, K is chosen as 10, and according to many experiments for more numbers we discover that number 10 is the one that gives the highest accuracy. In addition, we investigated several files of the subject and the majority advice to choose the number 10 to give better results and better learning ability. **cross validation works as follows:**

Divide the 10 groups into two parts as shown in the table above. The first section contains (9/10) groups which contains the part that the model will learn from, and the second section contains one set (1/10) which is the part that we will do a test on it and show us a certain accuracy, after that this process will make reverses groups that take 9 different groups for learning ⁶⁴

and one group for testing process and returns to calculate the accuracy in the same way and so on (the process does the switching process until it finishes all the groups (train, test) and takes the average of All the accuracy that appeared in all operations, and shown to us as a final result rate and this is the result on which to calculate whether the accuracy is high or not.

And this process contains sub processes as shown below:



Number 1 is the (training part) that contains the Classification model (decision tree)

Number 2 is the (testing part) that contains: -

1- The apply model (This Operator applies a model on an Example Set (train set).

2- The performance operator (This operator is used for performance evaluation. It delivers a list of performance criteria values. These performance criteria are automatically determined in order to fit the learning task type.)

After We click on RUN (\triangleright) we can notice the program run by see the cross validation operator

THE RESULT

accuracy: 83.06% +/- 5.56% (mikro: 83.08%)

	true negative	true positive	true neural	class precision
pred. negative	126	1	1	98.44%
pred. positive	0	103	2	98.10%
pred. neural	24	49	149	67.12%
class recall	84.00%	67.32%	98.03%	

Row No.	Result	prediction(R	confidence(confidence(confidence(aaroncart	abandon	abbygailram	ab	51
1	negative	negative	0.991	0	0.009	0	0	0	0	^
2	negative	negative	0.991	0	0.009	0	0	0	0	
3	negative	neural	0.077	0.699	0.224	0	0	0	0	
4	negative	neural	0.077	0.699	0.224	0	0	0	0	
5	negative	negative	0.991	0	0.009	0	0	0	0	
6	negative	negative	0.991	0	0.009	0	0	0	0	
7	negative	negative	0.991	0	0.009	0	0	0	0	
8	negative	negative	0.991	0	0.009	0	0	0	0	
9	negative	negative	0.991	0	0.009	0	0	0	0	
10	negative	negative	0.991	0	0.009	0	0	0	0	
11	negative	negative	0.991	0	0.009	0	0	0	0	
12	negative	negative	0.991	0	0.009	0	0	0	0	
13	negative	negative	0.991	0	0.009	0	0	0	0	
14	negative	negative	0.991	0	0.009	0	0	0	0	
15	negative	negative	0.991	0	0.009	0	0	0	0	~



Now we want to try a different classification model to see if the accuracy change (increase, decrease) to decide which classifier is better to use.

First we try by using decision tree

Description

```
hate > 0.034: negative {negative=126, positive=1, neural=0}
hate ≤ 0.034
| love > 0.068: positive {negative=0, positive=103, neural=1} |
love ≤ 0.068
| | fuck > 0.285: negative {negative=3, positive=0, neural=0}
| fuck ≤ 0.285
| | ass > 0.131: negative {negative=2, positive=0, neural=0} |
| ass ≤ 0.131
| | hit > 0.133: negative {negative=2, positive=0, neural=0}
| | hit ≤ 0.133: neural {negative=17, positive=49, neural=151}
```

68



Second we want to choose naïve bays classifier

The **confusion matrix** is as shown below:

🖲 Table View 🔘 Plot View

accuracy: 53.17% +/- 6.02% (mikro: 53.19%)

	true negative	true positive	true neural	class precision
pred. negative	99	50	46	50.77%
pred. positive	25	76	39	54.29%
pred. neural	26	27	67	55.83%
class recall	66.00%	49.67%	44.08%	

Description

Performance Vector: Accuracy: 53.17% +/- 6.02% (mikro: 53.19%) Confusion Matrix: negative positive neural True: Negative: 99 50 46 Positive: 25 76 39 neural: 262767kappa: 0.298 +/- 0.090 (mikro: 0.298) Confusion Matrix: negative positive neural True: negative: 99 50 46 positive: 25 7639 neural: 26 2767



Third we use SVM (support vector machine)

As we see the error, SVM does not support polynomial label (positive, negative, neutral).

Model	Accuracy
Decision Tree	83.06%
Naïve Bayes	53.17%
SVM	Does not support
Then we will let the machine execute what it learns from the training data we give for all data generated from the twitter about Walmart

First:

We load all twitter Walmart data into the Rapidminer using the read Excel process like shown below:



Second:

we use the Nominal to text process (To specify which column is a text column, since Rapidminer "Process Documents..." Operators work only on text data as shown below:



Third:

We use Process Documents from Data process that we used previously and we use the same sub process shown and described previously and they look like this:



And the sub processes are:



Forth:

We choose apply model operator to execute all above processes as shown below:



Problems that faces us in the project:

1. Tool not fully support some languages such as Arabic.

2. Tool does not support the huge number of data -

3. the existence of most comments in the colloquial language which is not analyzed by neither humans or tools.

4. tool must be paid to give better results in pulling data.

5. lack of sufficient resources to learn through.

6. The next step to work we were trying to find it with difficulty because the idea is new.

7. The difficulty of finding tool that performs the process of pulling the data fully and well.

Solutions & recommendations:

After reviewing and studying results, recommendations had to be made to solve negative problems related to the opinions of people on certain topics of Walmart:

1. they should pay more attention to the website and e-procurement because of many complaints related to Online Service.

2. there are also some customers who have complained about the bad treatment of the employees.

3. disturbing customers by the fact that the branches of Walmart have full of work pressure and crisis forcing customers to wait.

4.some complain that customers accuse Walmart of not paying attention of negative comments on social media and sometimes deleting them instead of dealing with them.

5.Online Chat does not work on their site as customers are not answered.6.We also recommend companies to do the sentiment analysis process in order to know their customers' opinions.

7.We also have the ability to analyze any company.

Thank you