

**LOB2203**  
**ระบบเทคโนโลยีสารสนเทศด้านโลจิสติกส์**  
**สำหรับธุรกิจออนไลน์**  
**CH 8-9**



สาขาการจัดการโลจิสติกส์สำหรับธุรกิจออนไลน์  
วิทยาลัยโลจิสติกส์และซัพพลายเชน มหาวิทยาลัยราชภัฏสวนสุนันทา



# Chapter 8

## Introduction to Text Mining





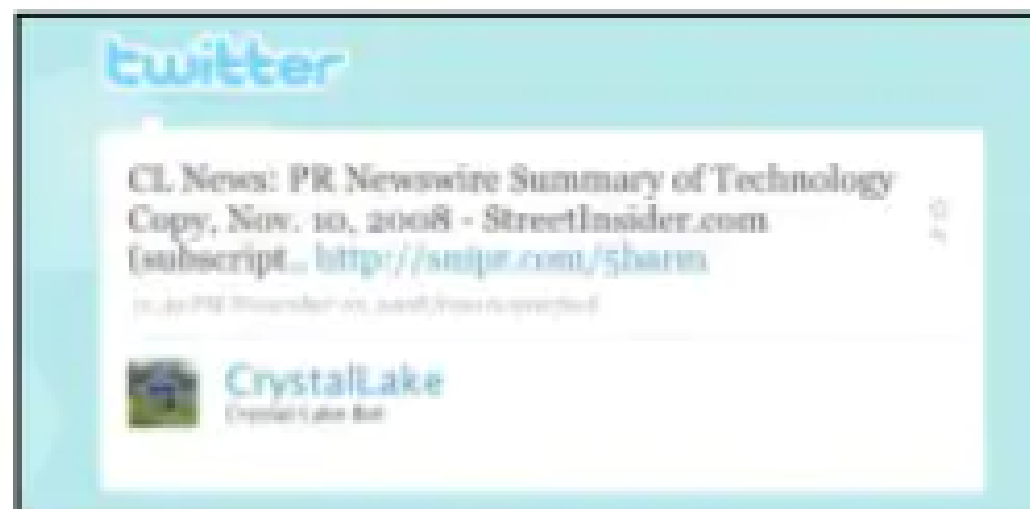
# Structured & Unstructured data

- ข้อมูลแบ่งเป็นแบบมีโครงสร้าง (structure) และไม่มีโครงสร้าง (unstructure)
- ข้อมูลแบบมีโครงสร้าง เช่น ข้อมูลที่เก็บในรูปแบบตาราง

ID	outlook	humidity	windy	play
1	sunny	high	FALSE	no
2	sunny	high	TRUE	no
3	overcast	normal	FALSE	yes

← numeric → ← nominal → ← binominal →

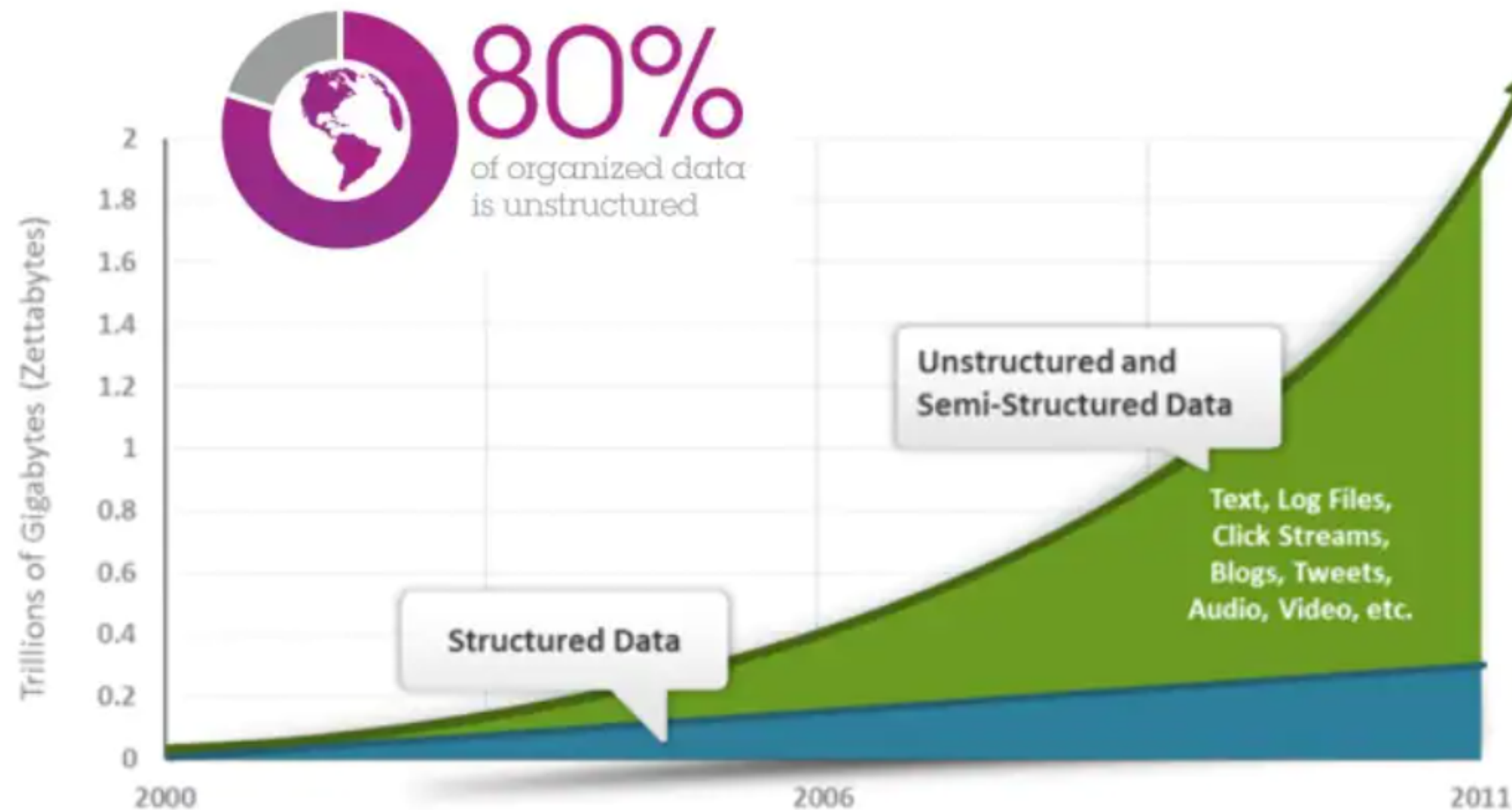
- ข้อมูลแบบไม่มีโครงสร้าง เช่น ข้อมูลที่เป็นข้อความ ข้อมูลที่เป็นรูปภาพ





## Structured & Unstructured data

- ข้อมูลที่ไม่มีโครงสร้างที่เก็บอยู่ในรูปแบบข้อความ รูปภาพ เสียงมีจำนวนมากถึง 80% ของข้อมูลทั้งหมด







## Why Text is important?

- ข้อมูลที่เป็นประเภทข้อความมีอยู่เป็นจำนวนมาก
  - ข้อความใน Social Network ต่างๆ เช่น Facebook, Twitter, LinkedIn
  - ข้อความในเอกสารต่างๆ เช่น email, SMS, รายงานต่างๆ
  - ข้อความในข่าวต่างๆ เช่น หนังสือพิมพ์, Google News

Leave Comments

10 comments [Add a comment](#)

**Jennifer Sable Lopez** - Community Manager at SEOmoz  
Looks like I'll be giving it another shot. :)  
Like - Reply - Subscribe - 13 minutes ago

**Rudy Lopez** - Seattle, Washington  
It looks like the iPhone app gets the upgrades first. The new desktop app is now just like it.  
Like - Reply - 4 minutes ago

**Janie Sawen** - VP Marketing at SEOmoz  
I've come to really love the client, just wish they'd change the ugly font  
Like - Reply - 58 seconds ago

Tech > Computer

### Appeal court revives Oracle-Google copyright battle

Published: 12 May 2014 at 08:48 (Viewed: 480) | Comments: 0  
Online news: Computer  
Writer: AFP

1 0  
Tweet 3-1  
Print Email Share

An appeals court has breathed new life into Oracle's big-money lawsuit against Google by ruling that software commands can be copyrighted just like classic books.



The case stems from 2012 trial, in which Oracle claimed Google owed them billions in damages for using parts of the Java programming language in its Android smartphone operating system.

The case is being closely watched in Silicon Valley, where some champions of internet freedom worry that extending copyright protection to these bits of code, called application programming interfaces, or APIs, would threaten innovation.

A panel of three judges in the US Federal Circuit Court of Appeals concluded that the trial court in 2012 erred and that it is bound to afford APIs protection under copyright law "unless either the Supreme Court or Congress tells us otherwise."

"We're disappointed -- and worried," the Electronic Frontier Foundation (EFF) said in a blog post about the appeals court.

**Bangkok Post Tech**



## Why text is important?

- การประยุกต์ใช้ข้อมูลประเภทข้อความ
  - หาข่าวที่ใกล้เคียง

The screenshot shows a Google News search result for 'WhatsApp'. The search bar at the top contains the word 'WhatsApp'. Below the search bar, the word 'News' is displayed in red. The main content area shows a list of news stories. The top story is titled 'Zuckerberg may have made a brilliant move with WhatsApp deal' and is from 'The Globe and Mail - 16 hours ago'. The second story is titled 'Mobile messenger WhatsApp to add voice calls' and is from 'The Globe and Mail - Feb 24, 2014'. The left sidebar contains a list of categories: Top Stories, News near you, World, Canada, Business, Technology, Samsung Group, WhatsApp, Nokia, BlackBerry, Sierra Wireless, Fraud, Ontario, Chevrolet Suburban, Google, and The internet. The bottom of the page shows a row of small thumbnail images from various news sources, including DigitalOut, The Province, Khaleej Times, Vancouver..., CNN, Toronto Star, MobileSynu..., and CTV News.



# Why text is important?

- การประยุกต์ใช้ข้อมูลประเภทข้อความ

Sentiment140

Tweet 741 Like 271 Retweet 175

iphone6

English

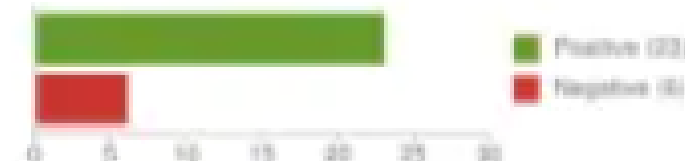
Search

Sentiment analysis for iphone6

Sentiment by Percent



Sentiment by Count



Tweets about: iphone6

**\_BigDaddyLiz:** RT @TheLifeSayin: The year is 2060. iPhone 842 is released. The screen touches you. #iPhone6

Posted: 21 seconds ago

**TokyoLife206:** RT @Syediant: finally #iphone6 is out. iPhone 5 will be cheaper now, so it means i can buy iPhone 4! :3

Posted: 24 seconds ago

**FahiraSuhairie:** RT @BellaShamsul: "@FahiraSuhairie: Too much drama on twitter bout iPhone6."

Posted: 28 seconds ago

The results for this query are:

Accurate

Inaccurate





## Text representation

- ในการวิเคราะห์ข้อมูลข้อความต้องทำการแปลงข้อมูลให้อยู่ในรูปแบบที่มีโครงสร้าง

**twitter**

Laying on the sand sun hitting me, this feels good :)

Smells so good I have two :)

i watched 3 movies today, reminded me of my child hood! Finding Nemo was always my favourite, but that was when i was MUCH younger! :)

I think NYC would be amazing, I really wanna go there :)

he never care,he never look at me..:(

i am so tired!! and my feet hurt :(

positive

negative





# Text representation: Bag of Words

- จำนวนความถี่ของคำที่เกิดขึ้นในข้อความ



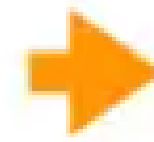




# Text representation: Bag of Words

- จำนวนความถี่ของคำที่เกิดขึ้นในข้อความ

ID	Tweet	Sentiment
1	Laying on the sand sun hitting me, this feels good.	positive
2	Smells so good I have two.	positive
3	i watched 3 movies today, reminded me of my child hood! Finding Nemo was always my favourite, but that was when i was MUCH younger!	positive
4	I think NYC would be amazing, I really wanna go there.	positive
5	he never care, he never look at me.	negative
6	i am so tired!! and my feet hurt.	negative



bag of words

Term	Count	Term	Count
Finding	1	child	1
I	3	favourite	1
Laying	1	feels	1
MUCH	1	feet	1
NYC	1	go	1
Nemo	1	good	2
Smells	1	have	1
always	1	he	2
am	1	hitting	1
amazing	1	hood	1
and	1	hurt	1
at	1	i	3
be	1	look	1
but	1	me	1
care	1	...	1



# Text representation: Stemming

- แปลงคำให้เป็นรากศัพท์ (root) เช่น finding แปลงเป็น find

Term	Count
Finding	1
I	3
Laying	1
MUCH	1
NYC	1
Nemo	1
Smells	1
always	1
am	1
amazing	1
and	1
at	1
be	1
but	1
care	1

Term	Count
find	1
i	6
lai	1
much	1
nyc	1
nemo	1
smell	1
alwai	1
am	1
amaz	1
and	1
at	1
be	1
but	1
care	1

Term	Count
child	1
favourite	3
feels	1
feet	1
go	1
good	2
have	1
he	2
hitting	1
hood	1
hurt	1
i	3
look	1
me	1
...	1

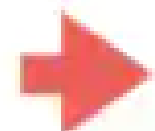
Term	Count
child	1
favourit	1
feel	1
feet	1
go	1
good	2
have	1
he	2
hit	1
hood	1
hurt	1
i	3
look	1
me	1
...	1



# Text representation: Remove stopwords

- ตัดคำที่เป็น stop word ทิ้ง

Term	Count	Term	Count
find	1	child	1
i	6	favourit	1
lai	1	feel	1
much	1	feet	1
nyc	1	go	1
nemo	1	good	2
smell	1	have	1
alwai	1	he	2
am	1	hit	1
amaz	1	hood	1
and	1	hurt	1
at	1		
be	1	look	1
but	1	me	1
care	1	...	1



remove stopwords

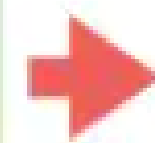
Term	Count	Term	Count
find	1	hood	1
i	6	hurt	1
lai	1	look	1
nyc	1	care	1
nemo	1	movi	1
smell	1	reali	1
alwai	1	remind	1
amaz	1	sand	1
child	1	sun	1
favorit	1	thi	1
feel	1	think	1
feet	1	tire	1
go	1	today	1
good	2	wa	3
hit	1	watch	1



# Convert Text to structured data

- Binary occurrence ถ้ามีคำที่เกิดขึ้นในแต่ละเอกสารจะให้ค่าเป็น 1  
 ถ้าไม่เกิดขึ้นจะเป็น 0

ID	Tweet	Sentiment
1	Laying on the sand sun hitting me, this feels good.	positive
2	Smells so good I have two.	positive
3	i watched 3 movies today, reminded me of my child hood! Finding Nemo was always my favourite, but that was when i was MUCH younger!	positive
4	I think NYC would be amazing, I really wanna go there.	positive
5	he never care,he never look at me.	negative
6	i am so tired!! and my feet hurt.	negative



ID	attribute								label
	find	I	lai	nyc	nemo	smell	alwai	...	Sentiment
1	0	0	1	0	0	0	0	...	positive
2	0	1	0	0	0	1	0	...	positive
3	1	1	0	0	1	0	1	...	positive
4	0	1	0	1	0	0	0	...	positive
5	0	0	0	0	0	0	0	...	negative
6	0	1	0	0	0	0	0	...	negative

training data แสดงค่าที่เกิดขึ้น ในเอกสารต่างๆ



# Convert Text to structured data

- Term Frequency คือจำนวนครั้งที่คำนั้นเกิดขึ้นในเอกสารทั้งหมด
  - TF = จำนวนคำในเอกสาร/จำนวนเอกสารทั้งหมด

ID	Tweet	Sentiment
1	Laying on the sand sun hitting me, this feels good.	positive
2	Smells so good I have two.	positive
3	i watched 3 movies today, reminded me of my child hood! Finding Nemo was always my favourite, but that was when i was MUCH younger!	positive
4	I think NYC would be amazing, I really wanna go there.	positive
5	he never care,he never look at me.	negative
6	i am so tired!! and my feet hurt.	negative



ID	attribute						label
	find	I	lai	nyc	nemo	...	Sentiment
1	0	0	0.17	0	0	...	positive
2	0	0.17	0	0	0	...	positive
3	0.17	0.33	0	0	0.17	...	positive
4	0	0.33	0	0.17	0	...	positive
5	0	0	0	0	0	...	negative
6	0	0.17	0	0	0	...	negative

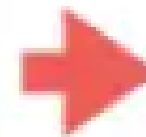
training data แสดงคำที่เกิดขึ้น ในเอกสารต่างๆ



# Convert Text to structured data

- TF-IDF คือจำนวนครั้งของคำที่เกิดขึ้นคูณกับจำนวนคำที่เกิดเฉพาะเอกสารในคลาส

ID	Tweet	Sentiment
1	Laying on the sand sun hitting me, this feels good.	positive
2	Smells so good I have two.	positive
3	i watched 3 movies today, reminded me of my child hood! Finding Nemo was always my favourite, but that was when i was MUCH younger!	positive
4	I think NYC would be amazing, I really wanna go there.	positive
5	he never care,he never look at me.	negative
6	i am so tired!! and my feet hurt.	negative



ID	attribute						label
	find	l	lai	nyc	nemo	...	Sentiment
1	0	0	0.40	0	0	...	positive
2	0	0.19	0	0	0	...	positive
3	0.22	0.10	0	0	0.22	...	positive
4	0	0.20	0	0.17	0	...	positive
5	0	0	0	0	0	...	negative
6	0	0.13	0	0	0	...	negative

training data แสดงคำที่เกิดขึ้นในเอกสารต่างๆ





## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## N-Gram sequences

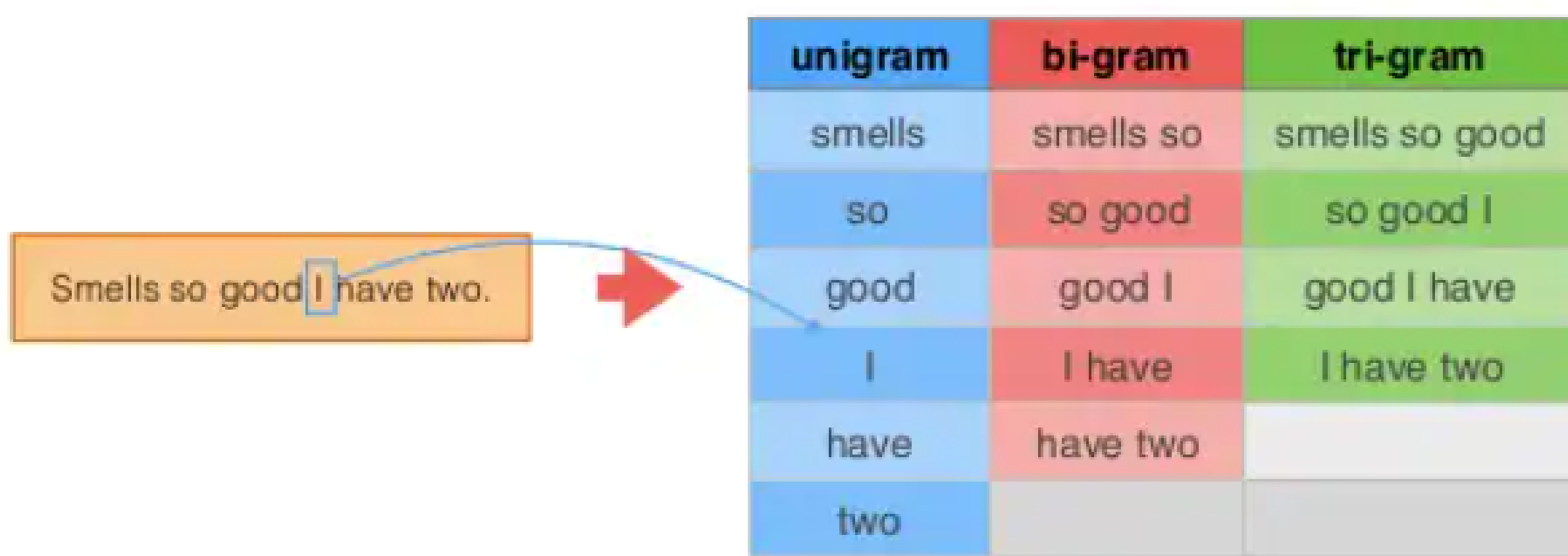
- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good





## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

Smells so good I have two.

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		





# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N คำ
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาดต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

Smells so **good** I have two.

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## N-Gram sequences

- จำนวนคำที่พิจารณาร่วมกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		





## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

Smells so good I have two.

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



# N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## N-Gram sequences

- จำนวนคำที่พิจารณาต่อกัน N ตัว
  - unigram พิจารณาการเกิดขึ้นของแต่ละคำ
  - bi-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 2 คำ เช่น so good
  - tri-gram พิจารณาการเกิดขึ้นของคำที่ติดกัน 3 คำ เช่น smells so good

unigram	bi-gram	tri-gram
smells	smells so	smells so good
so	so good	so good I
good	good I	good I have
I	I have	I have two
have	have two	
two		



## วิเคราะห์ข้อความ



rapidminer

# Text Mining



# วิเคราะห์ข้อความ

The screenshot shows the RapidMiner Studio Educational 9.9.002 @ TANAWAT interface. The title bar reads "<new process\*> - RapidMiner Studio Educational 9.9.002 @ TANAWAT". The menu bar includes File, Edit, Process, View, Connections, Settings, Extensions, and Help. The toolbar contains icons for file operations and a 'Views' dropdown menu with options: Design (selected), Results, Turbo Prep, Auto Model, Deployments, and More. On the left, the 'Repository' panel shows a tree structure with folders like Predictive Maintenance, Price Risk Clustering, Sentiment Analysis (containing Historical Sentiments and Sentiment Analysis), Web Analytics, and Time Series. The main 'Process' canvas shows a single step named 'Retrieve Historical Sentiments' connected to an input port 'inp' and an output port 'out'. The output port is connected to two 'res' ports on the right. The step icon features a green checkmark and a circular arrow.





# วิเคราะห์ข้อความ

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More Find

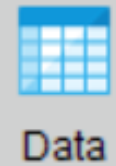
Result History

ExampleSet (Retrieve Historical Sentiments) X

Open in Turbo Prep Auto Model

Filter (200 / 200 examples): all

Row No.	Sentiment	Text
1	negative	" knock off " is exactly that : a cheap knock off of an action movie .
2	positive	ingredients : starving artist lusting after a beautiful woman from his childhood
3	negative	a remake of the 1978 warren beatty vehicle " heaven can wait " ( which in turn was a remake of " ...
4	positive	there's something about ben stiller that makes him a popular choice among casting directors th...
5	positive	note : some may consider portions of the following text to be spoilers .
6	positive	while watching wes anderson's rushmore , it may be surprising to think that the role of max fisch...
7	positive	dora ( fernanda montenegro ) sits behind a make-shift desk at rio+s major train station .
8	positive	well , i know that stallone is 50 years old now , but in daylight he doesn't look it !
9	negative	starting with the little mermaid and most recently the lion king , the walt disney company once ag...
10	negative	so , it's thirty years later , and oscar and felix are together again .



Data



Statistics



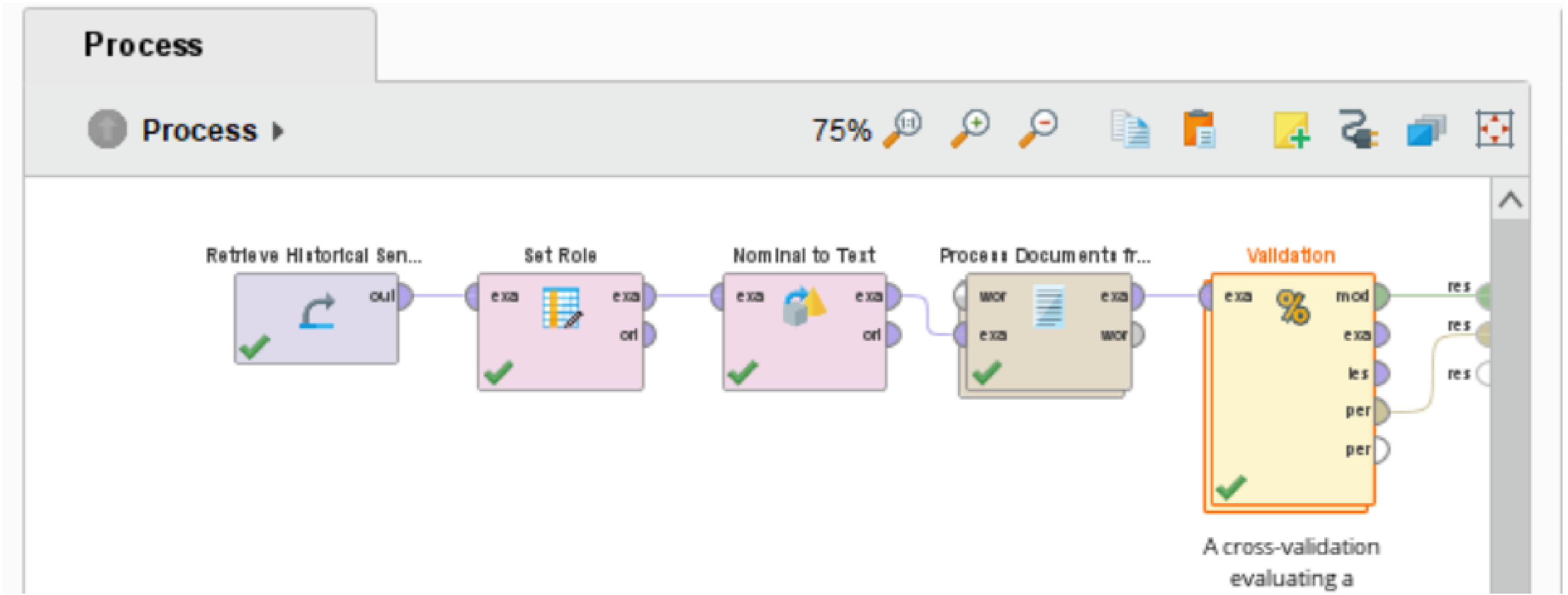
Visualizations



Annotations



# วิเคราะห์ข้อความ





# วิเคราะห์ข้อความ

The screenshot shows the RapidMiner Studio interface. The main workspace displays a workflow with two operators: 'Retrieve Historical ...' and 'Set Role'. The 'Set Role' operator is highlighted with an orange border. The 'Parameters' panel on the right shows the configuration for the 'Set Role' operator:

Parameters	
attribute name	Text
target role	regular
set additional roles	Edit List (0)...





# วิเคราะห์ข้อความ

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More

**Repository**

- Import Data
- Predictive Maintenance
- Price Risk Clustering
- Sentiment Analysis
  - Historical Sentiments
  - Sentiment Analysis
- Web Analytics

**Process**

90%

Retrieve Historical ... Set Role Nominal to Text

```
graph LR; A[Retrieve Historical ...] --> B[Set Role]; B --> C[Nominal to Text];
```



# วิเคราะห์ข้อความ

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More

Find data, operators...etc All Studio

**Repository**

- Import Data
- Predictive Maintenance
- Price Risk Clustering
- Sentiment Analysis
  - Historical Sentiments
  - Sentiment Analysis
- Web Analytics
- Time Series

**Process**

Process

Retrieve Historical S... out

Text Vectorization

inp

exa

exa

ori

pre

res

res

**Parameters**

Text Vectorization

attribute filter type: single

attribute: Text

invert selection

include special attributes

add sentiment

add language

keep original

store training documents

store scoring documents

document class attribute: Sentiment

Recommended Operators



# วิเคราะห์ข้อความ

Open in Turbo Prep Auto Model

Filter (200 / 200 examples):

Row No.	Sentiment	Text:table	Text:above	Text:absolut...	Text:academy	Text:accept	Text:accident	Text:acro
1	negative	0	0	0	0	0	0	0
2	positive	0	0	0	0	0	0	0
3	negative	0	0	0	0	0	0	0
4	positive	0	0	0	0	0	0	0
5	positive	0	0	0	0.003	0	0	0
6	positive	0	0	0	0.009	0	0	0
7	positive	0	0	0	0	0	0	0
8	positive	0	0	0	0	0	0	0
9	negative	0	0.008	0	0	0	0	0
10	negative	0	0	0	0	0	0	0
11	positive	0	0	0	0.006	0	0	0
12	positive	0.009	0	0	0	0	0	0
13	positive	0	0	0	0	0	0	0



# วิเคราะห์ข้อความ

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

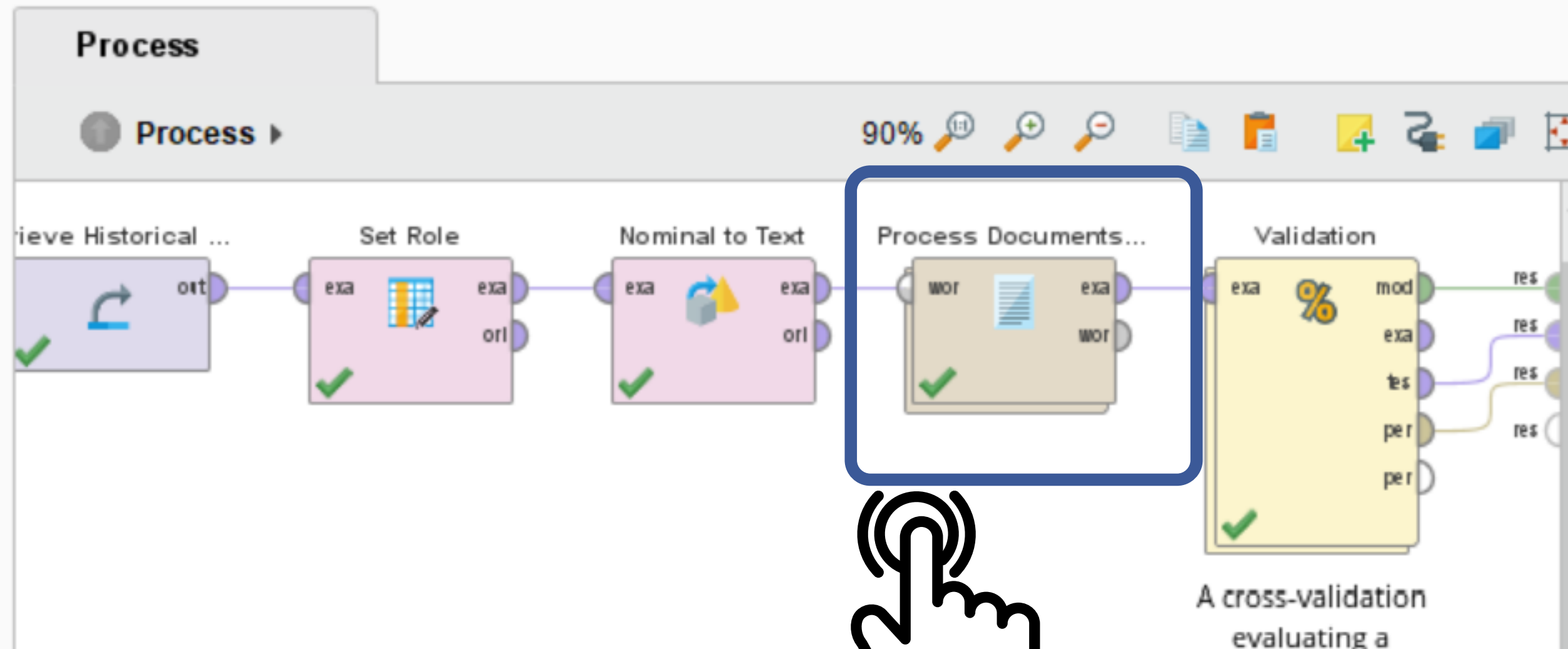
File Edit Process View Connections Settings Extensions Help



**Repository**

Import Data

- Text Processing
  - Apply Model Documents
  - Dictionary Sentiment Learner
  - Filter Tokens from ExampleS
    - Filter Tokens from Examp
  - LDA
  - Split Document into Collectio





# วิเคราะห์ข้อความ

The screenshot displays the RapidMiner Studio Educational 9.9.002 @ TANAWAT interface. The main window is titled "<new process\*> - RapidMiner Studio Educational 9.9.002 @ TANAWAT". The menu bar includes File, Edit, Process, View, Connections, Settings, Extensions, and Help. The toolbar contains icons for file operations and a play button. The Views section shows Design, Results, Turbo Prep, Auto Model, Deployments, and More. A search bar contains "Find data, operators...etc" and "All Studio".

The interface is divided into several panels:

- Repository:** Shows a tree view of data sources and processes. The "Sentiment Analysis" folder is expanded, showing "Historical Sentiments" and "Sentiment Analysis".
- Process:** Shows a workflow diagram with a single operator "Process Documents from Data" connected to a data source "doc".
- Parameters:** Shows the configuration for the "Process Documents from Data" operator. The parameters are:
  - create word vector
  - vector creation: TF-IDF
  - add meta information
  - keep text
  - prune method: none
  - data management: auto
  - select attributes and weights
- Operators:** Shows a search bar with "process da" entered.





# วิเคราะห์ข้อความ

The screenshot shows the RapidMiner Studio interface. The main workspace displays a process flow: 'Process Documents from Files' -> 'Filter Tokens (by Length)' -> 'Stem (Porter)' -> 'Filter Stopwords (En...)'. The 'Filter Tokens (by Length)' operator is selected, and its parameters are visible in the right-hand panel:

- min chars: 4
- max chars: 25

- Filter Token (by Length) = สำหรับกรองคำศัพท์ที่มีความยาวน้อยกว่าหรือมากกว่าที่กำหนด
- Stem (Porter) = สำหรับแปลงคำให้อยู่ในรูปรากศัพท์
- Filter Stopwords = สำหรับลบคำที่เป็น Stopwords เช่น a,an,the



# วิเคราะห์ข้อความ

<new process\*> - RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More

**Repository**

- Import Data
- Predictive Maintenance
- Price Risk Clustering
- Sentiment Analysis
  - Historical Sentiments
  - Sentiment Analysis
- Web Analytics
- Time Series

**Process**

75%

Retrieve Historical Sen... Set Role Nominal to Text Process Documents fr... Validation

A cross-validation evaluating a decision tree model.



# วิเคราะห์ข้อความ

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More

Result History PerformanceVector (Performance) Tree (Decision Tree)

Performance

Criterion

- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Description

Table View Plot View

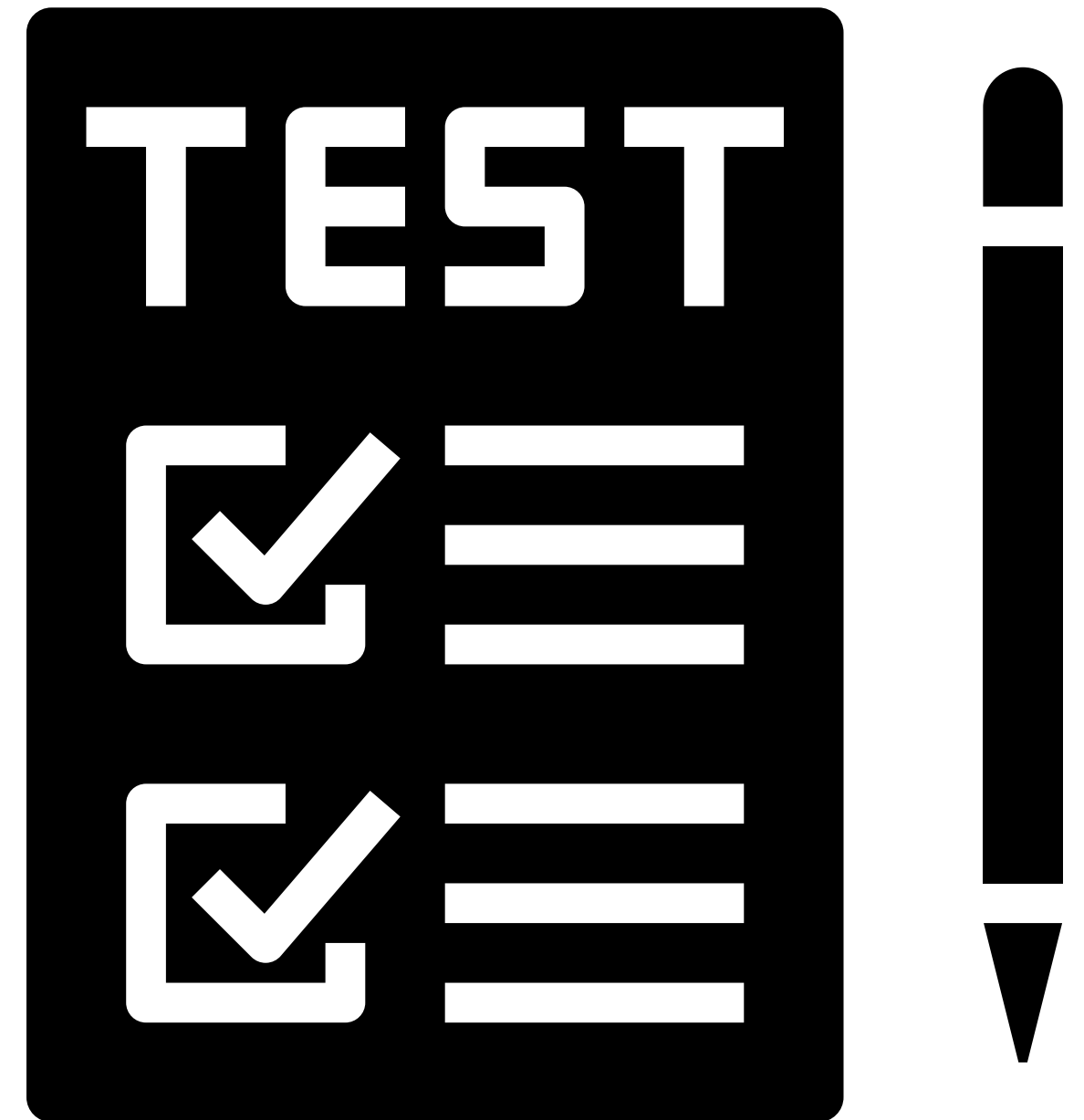
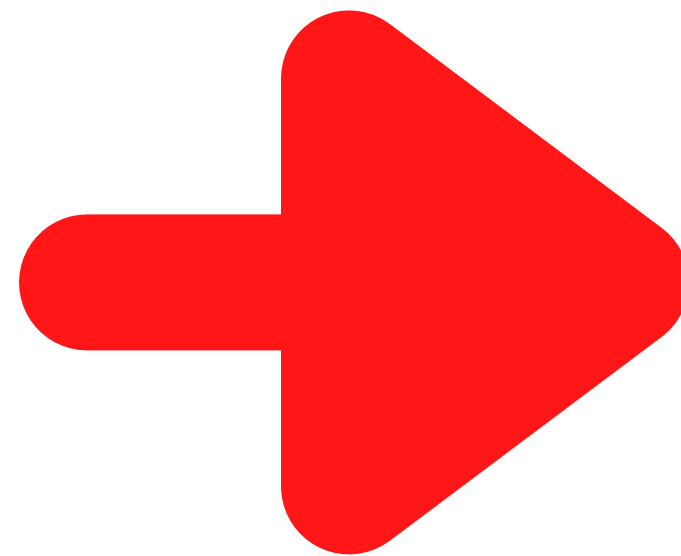
accuracy: 52.50% +/- 2.64% (micro average: 52.50%)

	true negative	true positive	class precision
pred. negative	0	0	0.00%
pred. positive	95	105	52.50%
class recall	0.00%	100.00%	



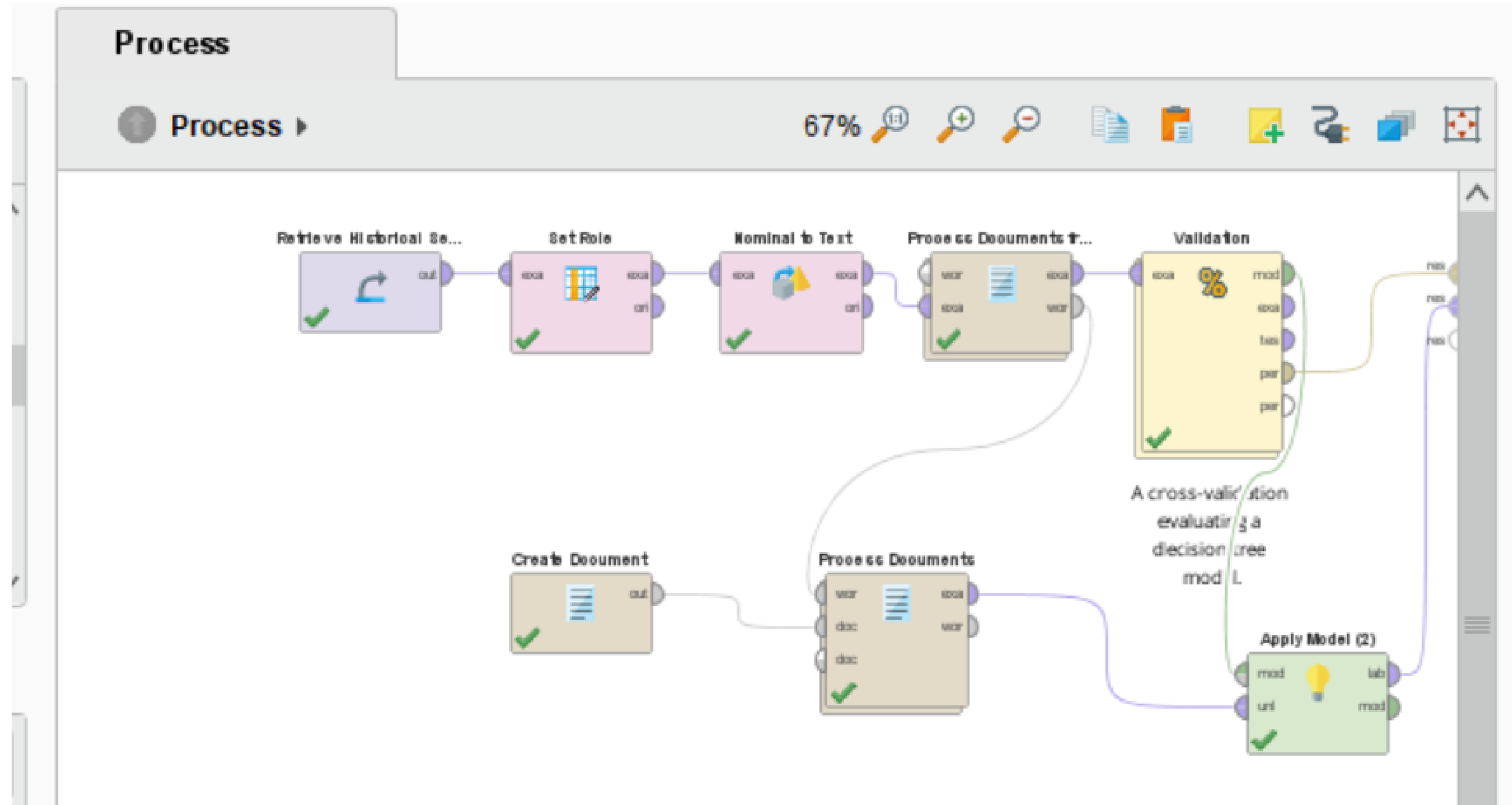


## นำข้อมูลมาทดลองจริง





# นำข้อมูลมาทดลองจริง





# นำข้อมูลมาทดลองจริง

The screenshot displays the RapidMiner Studio Educational 9.9.002 interface. The main workspace shows a workflow with the following operators: 'out' (input), 'exa' (Excel Reader), 'ori' (Orientation), 'exa' (Excel Writer), 'wor' (Word Writer), and 'exa' (Excel Writer). A 'Create Document' operator is highlighted with an orange box. The 'Parameters' panel on the right shows the 'Create Document' operator with the 'text' parameter set to 'i love you|'. The 'Edit Parameter Text: text' dialog box is open, showing the text 'The text.' and the input '1 i love you|'.



# นำข้อมูลมาทดลองจริง

<new process\*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More app

Result History ExampleSet (Apply Model (2)) PerformanceVector (Performance)

Open in Turbo Prep Auto Model Filter (1 / 1 examples): all

Row No.	prediction(Sentiment)	confidence(negative)	confidence(positive)	" crazy/bea...	" desperate...	"
1	positive	0.475	0.525	0	0	0