

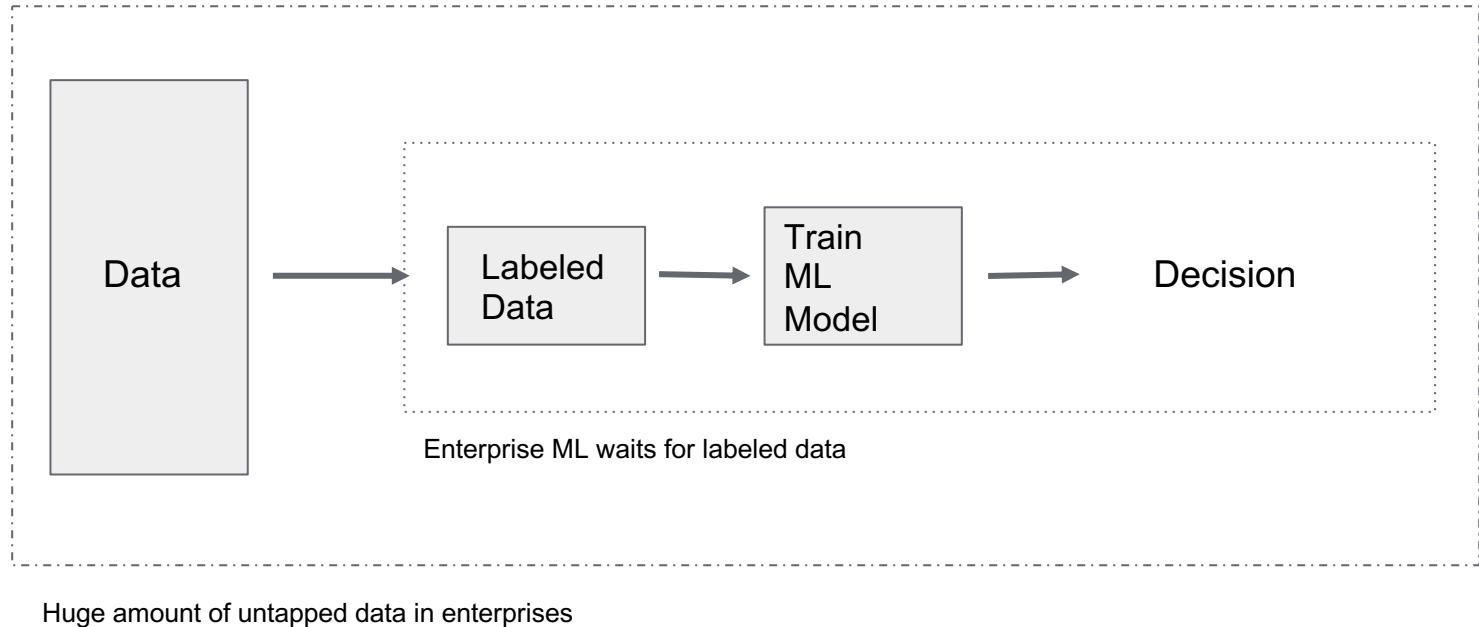
Snorkel

OLD VERSION _ DO NOT USE

The Mechanics of Data-First AI

- Rajiv Shah
- raj@snorkel.ai

What is limiting AI in Enterprises?



Someone has to label the data

01/01/201X - Pod 1 AM

Room #: 01

Enc. #: 01

Routing #: 1

NBOME ID #: 111111

S

George Payne is a 45-year-old Caucasian right handed male who presents to the Family Medicine office with a complaint of right chest discomfort for one week. The pain began after moving into a new house. He denies any injury to his chest. The pain is sharp, constant, and he rates it 3-4/10. The pain radiates through to the back intermittently. It is relieved with a hot shower, ibuprofen, and rest. He has never had anything like this previously. He is concerned that this could be related to his heart.

ROS: No palpitations, shortness of breath, diaphoresis, nausea, or vomiting

Pmhx: heartburn occasionally

Surghx: tonsillectomy

Meds: Ibuprofen three times a day

ALL: NKDA

Fam Hx: Mother living and in good health

Sochx: Married with 3 children. L

Vitals: 70 inches 190 lbs 27.3

Gen: mildly anxious male in mild distress

Heart: regular, no murmurs, 53 BPM

Lung: clear bilateral anterior and posterior

T spine: Paravertebral tenderness

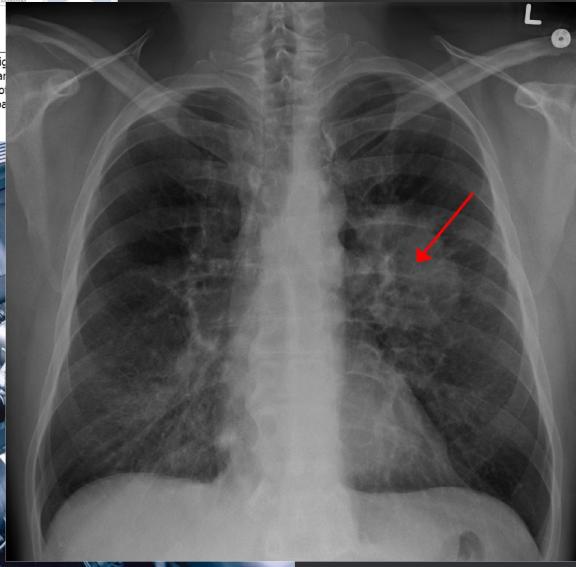
Chest wall: Tenderness with palpation

A

1. Thoracic somatic dysfunction
2. Costochondritis
3. Rib fracture-unlikely
4. Anginal equivalent-doubt
5. Family history of heart disease



SOAP Note

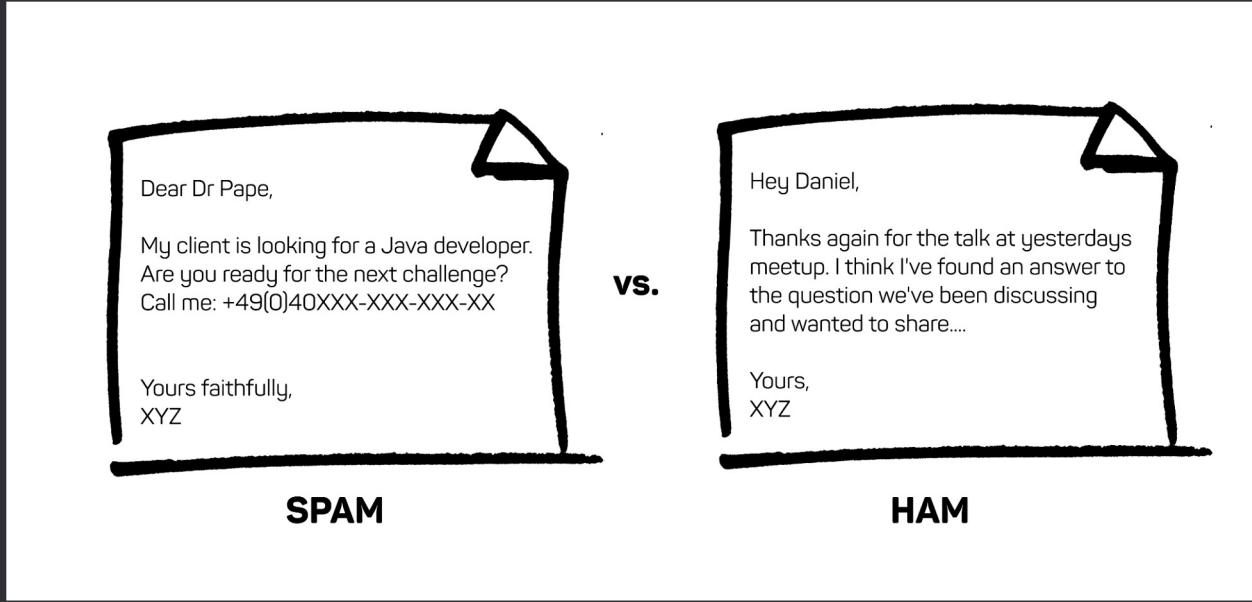


Issues:

Experts
Access to the Data
Lots of time
Opaque
Inflexible

Programmatic Labeling with Snorkel

Classifying messages as Spam



To build a machine learning model to classify between spam and real message, you provide many examples of spam and not spam (ham) to the computer.

A photograph of three people in an office setting, all appearing stressed or tired. A woman in a white blazer on the left has her hand to her chin. In the center, a man in a light blue shirt has his hands to his face. On the right, a woman in a blue blouse also has her hands to her head. They are seated around a white desk with a laptop, papers, and office supplies like pens and glasses. The lighting is warm and focused on the desk area.

Let's go label the data

a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more about this amazing investment opportunity and how a small investment could mean huge gains for you ! there is no doubt that china stocks , which are new to u . s . stock markets , are destined to blast off . it happens time and time again . thats why informed investors like warren buffett are getting

A	B	C	F
Message ID	Subject	Message	Spam/Han
26749	hot stock info	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	
28231	bad proposed	attached below is the cpuc 's proposed decision in the gas industryrestruc	
5619	material for	fyi . - - - - forwarded by ravi thuraisingham / enron communications on 03	
446	bpa rate case	- - - - - forwarded by mary hain / hou / ect on 03	
10684	enjoy media	enjoy media (ejym) , a chinese media company , on fast growth track by	
30380	next of kin [goodday , i am frenkel heydorn , staff of smith came to our bank to engag	
31489	are you a per	breaking news alertdigisound inc . (nqb : dgsu)music rights verification ,	
5617	[amy @ sds	vince , stinson , looks like just me and jim irvine attending this one mays	
7144	energy finan	david ,here are the reviews of the ut energy finance program that you rec	
28462	fw : california	alertyfy! - - - original message - - - from : webmaster @ cera . com @	
28285	california on	- - - - - forwarded by lorna brennan / et & s / enr	
11555	california up	executive summary? davis struggles to find senate support for socal tran	
4250	breaking new	daily traders reportnxfahxdbreaking newsvhxcwelefsqwle - directto low	
3533	re : path mai	since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that	
18583	news alert (what is tmxt technologies ?tmxt issued 2 news releases today , one durin	
28416	fw : western	latest western markets forecast from cambridge energy .lorna- - - or	
20191	paliourg , lor	brand new acquisition , reverse merger , ground floor !start watching thi	
5677	gt symposium	please share the following announcement with your associates .georgia	
3945	ad - when he	to their first time / titlemeta http - equiv = content - type content = text	
14151	an article fro	this story was sent to you by peggy . mccurley @ enron . com with these	
20044	re : my , wat	mona vang ,mineral exploration stockdelta mining and exploration corp .	
8785	re : summer	thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify v	
11091	davis claims	fyi .in the attached article , gov davis mentions that enron is one of the g	

a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more about this amazing investment opportunity and how a small investment could mean huge gains for you ! there is no doubt that china stocks , which are new to u . s . stock markets , are destined to blast off . it happens time and time again . thats why informed investors like warren buffett are getting

A	B	C	F
Message ID	Subject	Message	Spam/Han
26749	hot stock info	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	SPAM
28231	bad proposal	attached below is the cpuc 's proposed decision in the gas industry restructuring	
5619	material for fyi	----- forwarded by ravi thuraisingham / enron communications on (
446	bpa rate cast	----- forwarded by mary hain / hou / ect on 03	
10684	enjoy media	enjoy media (ejym) , a chinese media company , on fast growth track by	
30380	next of kin [goodday , i am frenkel heydorn , staff of smith came to our bank to engage	
31489	are you a per	breaking news alertdigisound inc . (nqb : dgsu)music rights verification ,	
5617	[amy @ sds	vince , stinson , looks like just me and jim irvine attending this one maybe	
7144	energy finan	david ,here are the reviews of the ut energy finance program that you recd	
28462	fw : california alert	fyi ----- original message ----- from : webmaster @ cera . com @	
28285	california on	----- forwarded by lorna brennan / et & s / enr	
11555	california up	executive summary? davis struggles to find senate support for socal trans	
4250	breaking news	daily traders reportnxfahxdbreaking newsvhxcwevlefsqwl - direct to low	
3533	re : path mail	since lisa used fuchsia , i ' ll use lavender . on # 1 i would eliminate that	
18583	news alert (what is tmxt technologies ?tmxt issued 2 news releases today , one durin	
28416	fw : western	latest western markets forecast from cambridge energy .lorna----- or	
20191	paliourg , lor	brand new acquisition , reverse merger , ground floor !start watching thi	
5677	gt symposium	please share the following announcement with your associates .georgia ----- or	
3945	ad - when he	to their first time / titlemeta http - equiv = content - type content = text ,	
14151	an article fro	this story was sent to you by peggy . mccurley @ enron . com with these	
20044	re : my , wat	mona vang ,mineral exploration stockdelta mining and exploration corp .	
8785	re : summer	thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify v	
11091	davis claims	fyi .in the attached article , gov davis mentions that enron is one of the g	

attached below is the cpuc ' s proposed decision in the gas industry ▲
restructuring proceeding . essentially , it adopts the interim
proposal with a
few modifications . the cpuc states that they are taking a "
cautious "
approach to deregulation in light of recent market events in both
electricity
and gas , and therefore , rejected the more robust reforms
outlined in the

A	B	C	F
Message ID	Subject	Message	Spam/Ham
26749	hot stock info	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	SPAM
28231	bad propose	attached below is the cpuc ' s proposed decision in the gas industry restruc	
5619	material for	fyi . - - - - forwarded by ravi thuraisingham / enron communications on (
446	bpa rate cas	----- forwarded by mary hain / hou / ect on 03	
10684	enjoy media	enjoy media (ejym) , a chinese media company , on fast growth track b	
30380	next of kin [goodday , i am frenkel heydorn , staff of smith came to our bank to engag	
31489	are you a per	breaking news alertdigsound inc . (nqb : dgsu)music rights verification ,	
5617	[amy @ sds	vince , stinson , looks like just me and jim irvine attending this one mays	
7144	energy finan	david ,here are the reviews of the ut energy finance program that you rec	
28462	fw : californi	alertyfi - - - original message - - - -from : webmaster @ cera . com @	
28285	california on	----- forwarded by lorna brennan / et & s / enr	
11555	california up	executive summary? davis struggles to find senate support for so cal tran	
4250	breaking new	daily traders reportnxfahxdbreaking newsvhxewelevfsqle - directto low	
3533	re : path mai	since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that	
18583	news alert (what is tmxt technologies ?tmxt issued 2 news releases today , one duri	
28416	fw : western	latest western markets forecast from cambridge energy .lorna- - - or	
20191	paliourg , lor	brand new acquisition , reverse merger , ground floor !start watching thi	
5677	gt symposium	please share the following announcement with your associates .georgia	
3945	ad - when	he to their first time / titlemeta http - equiv = content - type content = text ,	
14151	an article fro	this story was sent to you by peggy . mccurley @ enron . com with these	
20044	re : my , wat	mona vang ,mineral exploration stockdelta mining and exploration corp .	
8785	re : summer	thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify v	
11091	davis claims	fyi .in the attached article , gov davis mentions that enron is one of the g	

▲ X ✓ fx

attached below is the cpuc 's proposed decision in the gas industry ▲
restructuring proceeding . essentially , it adopts the interim
proposal with a
few modifications . the cpuc states that they are taking a "cautious"
approach to deregulation in light of recent market events in both
electricity
and gas , and therefore , rejected the more robust reforms
outlined in the

A	B	C	F
Message ID	Subject	Message	Spam/Ham
26749	hot stock info a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	SPAM	
28231	bad propose attached below is the cpuc 's proposed decision in the gas industryrestrict	HAM	
5619	material for fyi . - - - - forwarded by ravi thuraisingham / enron communications on 03		
446	bpa rate case . - - - - - forwarded by mary hain / hou / ect on 03		
10684	enjoy media enjoy media (ejym) , a chinese media company , on fast growth track by		
30380	next of kin [goodday , i am frenkel heydorn , staff of smith came to our bank to engag		
31489	are you a per breaking news alertdigisound inc . (nqb : dgsu)music rights verification ,		
5617	[amy @ sds vince , stinson , looks like just me and jim irvine attending this one mays		
7144	energy finan david ,here are the reviews of the ut energy finance program that you rec		
28462	fw : californi alertfyi- - - original message - - - -from : webmaster @ cera . com @		
28285	california on - - - - - forwarded by lorna brennan / et & s / enr		
11555	california up executive summary? davis struggles to find senate support for socal tran		
4250	breaking new daily traders reportnxfahxdbreaking newsvhxcwevlefsqwle - directto low		
3533	re : path mai since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that		
18583	news alert (what is tmxt technologies ?tmxt issued 2 news releases today , one duri		
28416	fw : western latest western markets forecast from cambridge energy.lorna- - - or		
20191	paliourg , lor brand new acquisition , reverse merger , ground floor lstart watching thi		
5677	gt symposium please share the following announcement with your associates .georgia		
3945	ad - when he to their first time / titlemeta http - equiv = content - type content = text ,		
14151	an article fro this story was sent to you by peggy . mcurley @ enron . com with these		
20044	re : my , wat mona vang ,mineral exploration stockdelta mining and exploration corp .		
8785	re : summer thanks dr . kaminski ,i forgot that : -) i ' ll understand it asi may identify v		
11091	davis claims fyi .in the attached article , gov davis mentions that enron is one of the g		

▲ X ✓ jx

enjoy media (ejym) , a chinese media company , on fast growth
track by signing up chinese advertising clients
enjoy media anticipates strong growth in 2005
jun 7 , 2005 6 : 00 : 00 am
copyright business wire 2005
hong kong - - (business wire) - - june 7 , 2005 - -
enjoy media holdings limited (pink sheets : ejym) (www .
enjoymedia . com) , a print media and advertising services
company in china , discusses about the company ' s objectives and

A	B	C	F
Message ID	Subject	Message	Spam/Ham
26749	hot stock info	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	SPAM
28231	bad propose	attached below is the cpuc ' s proposed decision in the gas industry	HAM
5619	material for fyi	----- forwarded by ravi thuraisingham / enron communications on	HAM
446	bpa rate cas	----- forwarded by mary hain / hou / ect on 03	HAM
10684	enjoy media	enjoy media (ejym) , a chinese media company , on fast growth track b	
30380	next of kin	[goodday , i am frenkel heydorn , staff of smith came to our bank to engage	
31489	are you a per	breaking news alertdigisound inc . (nqb : dgsu)music rights verification ,	
5617	[amy @ sds	vince , stinson , looks like just me and jim irvine attending this one mays	
7144	energy finan	david ,here are the reviews of the ut energy finance program that you rec	
28462	fw : californi	alertyi ----- original message -----from : webmaster @ cera . com @	
28285	california on	----- forwarded by lorna brennan / et & s / enr	
11555	california up	executive summary? davis struggles to find senate support for socal tran	
4250	breaking new	daily traders reportnxfahxdbreaking newsvhxcwevlefsqwe - directto low	
3533	re : path mai	since lisa used fuchsia , i ' ll use lavender ..on # 1 i would eliminate that	
18583	news alert	(what is tmxt technologies ?tmxt issued 2 news releases today , one durin	
28416	fw : western	latest western markets forecast from cambridge energy .lorna----- or	
20191	paliourg	, lor brand new acquisition , reverse merger , ground floor Istart watching thi	
5677	gt symposium	please share the following announcement with your associates .georgia	
3945	ad - when he	to their first time / titlemeta http - equiv = content - type content = text ,	
14151	an article fro	this story was sent to you by peggy . mccurley @ enron . com with these	
20044	re : my , wat	mona vang ,mineral exploration stockdelta mining and exploration corp .	
8785	re : summer	thanks dr . kaminski ,i forgot that :-)i ' ll understand it asi may identify v	
11091	davis claims	fyi .in the attached article , gov davis mentions that enron is one of the g	

enjoy media (ejym) , a chinese media company , on fast growth
 track by signing up chinese advertising clients
 enjoy media anticipates strong growth in 2005
 jun 7 , 2005 6 : 00 : 00 am
 copyright business wire 2005
 hong kong -- (business wire) -- june 7 , 2005 --
 enjoy media holdings limited (pink sheets : ejym) (www .
 enjoymedia . com) , a print media and advertising services
 company in china , discusses about the company ' s objectives and

A	B	C	D
Message ID	Subject	Message	Spam/Ham
26749	hot stock info a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time	SPAM	
28231	bad proposal attached below is the cpuc ' s proposed decision in the gas industryres	ru	HAM
5619	material for fyi - - - - forwarded by ravi thuraisingham / enron communications on	(HAM
446	bpa rate case - - - - - forwarded by mary hain / hou / ect on 03)	HAM
10684	enjoy media	enjoy media (ejym) , a chinese media company , on fast growth track b	SPAM
30380	next of kin [goodday , i am frenkel heydorn , staff of smith came to our bank to engage		
31489	are you a per breaking news alertdigisound inc . (nqb : dgsu)music rights verification ,		
5617	[amy @ sds vince , stinson , looks like just me and jim irvine attending this one mays		
7144	energy finan david ,here are the reviews of the ut energy finance program that you rec		
28462	fw : californi alertfy! - - - original message -----from : webmaster @ cera . com @		
28285	california on - - - - - forwarded by lorna brennan / et & s / enr		
11555	california up executive summary? davis struggles to find senate support for socal tran		
4250	breaking new daily traders reportnxfahxdbreaking newsvhxcwevlefsqwl - directto low		
3533	re : path mai since lisa used fuchsia , i ' ll use lavender ..on # 1 i would eliminate that		
18583	news alert (what is tmxt technologies ?tmxt issued 2 news releases today , one durin		
28416	fw : western latest western markets forecast from cambridge energy .lorna- - - - or		
20191	paliourg , lor brand new acquisition , reverse merger , ground floor !start watching thi		
5677	gt symposium please share the following announcement with your associates .georgia		
3945	ad - when he to their first time / titlemeta http - equiv = content - type content = text ,		
14151	an article fro this story was sent to you by peggy . mcurley @ enron . com with these		
20044	re : my , wat mona vang ,mineral exploration stockdelta mining and exploration corp .		
8785	re : summer thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify v		
11091	davis claims fyi .in the attached article , gov davis mentions that enron is one of the g		



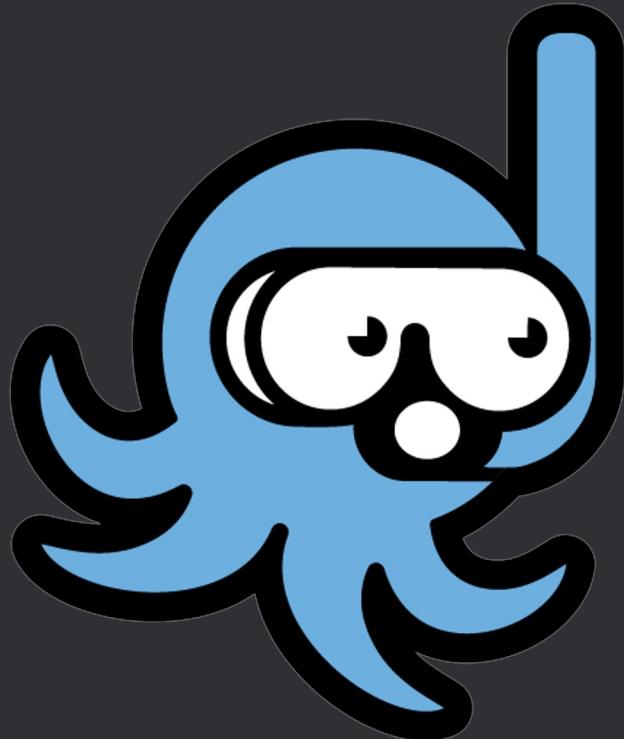
A TON OF
SCHOOL DAYS
LATER...

Message

Spam/Ha

a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time SPAM
attached below is the cpuc ' s proposed decision in the gas industryrestu HAM
fyi . - - - - forwarded by ravi thuraisingham / enron communications on (HAM
----- forwarded by mary hain / hou / ect on 03 HAM
enjoy media (ejym) , a chinese media company , on fast growth track ! SPAM
goodday , i am frenkel heydorn , staff of smith came to our bank to engag SPAM
breaking news alertdigsound inc . (nqb : dgsu)music rights verification , SPAM
vince , stinson , looks like just me and jim irvine attending this one may HAM
david ,here are the reviews of the ut energy finance program that you rec HAM
alertfyi- - - - original message - - - -from : webmaster @ cera . com @ HAM
----- forwarded by lorna brennan / et & s / enr HAM
executive summary? davis struggles to find senate support for socal train HAM
daily traders reportnxfahxdbreaking newsvhxcwevlefsqwle - directto lov SPAM
since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that HAM
what is tmxt technologies ?tmxt issued 2 news releases today , one dur SPAM
latest western markets forecast from cambridge energy .lorna- - - - or HAM
brand new acquisition , reverse merger , ground floor !start watching thi SPAM
please share the following announcement with your associates .georgia HAM
to their first time / titlemeta http - equiv = content - type content = text , SPAM
this story was sent to you by peggy . mccurley @ enron . com with these HAM
mona vang ,mineral exploration stockdelta mining and exploration corp . SPAM
thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify v HAM
fyi .in the attached article , gov davis mentions that enron is one of the g HAM
it was the spring of 1979 .i was just a tall , goofy looking kid in middle sc SPAM
cera alert : december 13 , 2000title : california on the brinkcera knowledg HAM
we prepared a presentaton on argentina and brazil for your sp trip . mich HAM
hello shmuwel ,thanks for your message . the end of 2000 and the beginni HAM
jeff ,we shall forward to you shortly a copy of the message from sandeep HAM
the stock to buylearn what jp morgan , sg cowan and the wall street jour SPAM
mineral exploration stockdelta mining and exploration corp . otc : dmxp SPAM

Programmatic Labeling



a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more about this amazing investment opportunity and how a s
 you !
 there
 is no doubt that china stocks , which are new to u . s . stock markets , are destined to blast off . it happens time and time and time
 again . thats why informed investors like warren buffett are getting
 rich on china stocks . the market is enormous and now its your
 turn . the upside potential for drgv is huge . with potential revenues of nearly

B	C	G
Subject	Message	Spam/Ham
hot stock info : drgv position	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more	
bad proposed decision	attached below is the cpuc ' s proposed decision in the gas industry	
material for thu 2 mar wg rfyi .		
bpa rate case - seasonal rat-----	forwarded by mary hain / hou / ect on 03 / 14 / 2000 08 :	
enjoy media (ejym) enters	enjoy media (ejym) , a chinese media company , on fast growth track by signing up	
next of kin [business]	goodday ,	
are you a penny stox player breaking news alert		
[amy @ sdsc . edu : re : caivince , stinson , looks like just me and jim irvine attending this one may		
energy finance critiques	david ,	
fw : california ' s desperate alert		
california on the brink -- ce-----	forwarded by lorna brennan / et & s / enron on 12 / 14 /	
california update 4 / 20 / 01 executive summary		
breaking news : abdv goes daily traders report		
re : path manager rewrite / since lisa used fuchsia , i ' ll use lavender ..		
news alert (tmxt) receives what is tmxt technologies ?		
fw : western energy market latest western markets forecast from cambridge energy .		
paliourg , lorenzo 98 % of o brand new acquisition , reverse merger , ground floor !		
gt symposium on qcf , april please share the following announcement with your associates .		
ad - when he said he could to their first time / title		
an article from cbs . market this story was sent to you by peggy . mccurley @ enron . com with these comments :		
re : my , water fell down mona vang ,		

Programmatic thinking:

What are the steps for labeling messages?

a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more about this amazing investment opportunity and how a s
 you !
 there
 is no doubt that china stocks , which are new to u . s . stock markets , are destined to blast off . it happens time and time and time
 again . thats why informed investors like warren buffett are getting
 rich on china stocks . the market is enormous and now its your
 turn . the upside potential for drgv is huge . with potential revenues of nearly

B	C	G
Subject	Message	Spam/Ham
9 hot stock info : drgv position	a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more	
1 bad proposed decision	attached below is the cpuc ' s proposed decision in the gas industry	
9 material for thu 2 mar wg rfyi .		
6 bpa rate case - seasonal rat-----	forwarded by mary hain / hou / ect on 03 / 14 / 2000 08 :	
4 enjoy media (ejym) enters	enjoy media (ejym) , a chinese media company , on fast growth track by signing up	
0 next of kin [business]	goodday ,	
9 are you a penny stox player breaking news alert		
.7 [amy @ sdsc . edu : re : caivince , stinson , looks like just me and jim irvine attending this one may		
4 energy finance critiques	david ,	
2 fw : california ' s desperate alert		
5 california on the brink -- ce-----	forwarded by lorna brennan / et & s / enron on 12 / 14 /	
5 california update 4 / 20 / 01 executive summary		
0 breaking news : abdv goes e daily traders report		
3 re : path manager rewrite / since lisa used fuchsia , i ' ll use lavender ..		
3 news alert (tmxt) receives what is tmxt technologies ?		
6 fw : western energy market latest western markets forecast from cambridge energy .		
1 paliorug , lorenzo 98 % of o brand new acquisition , reverse merger , ground floor !		
'7 gt symposium on qcf , april	please share the following announcement with your associates .	
5 ad - when he said he could	to their first time / title	
1 an article from cbs . market	this story was sent to you by peggy . mccurley @ enron . com with these comments :	
4 re : my , water fell down	mona vang ,	

- Keywords
 - Investments
 - "Get rich"
 - "As seen on"
- If From Email Matches Address Book

Getting Programmatic

- Keywords
 - Investments
 - “Get rich”
 - “As seen on”
- If From Email Matches Address Book

Concepts



```
if Keyword in message:  
    label(SPAM)
```

```
if MatchEmail=True:  
    label(HAM)
```

Formula / Code

Message	Investments
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring proce fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0	SPAM
----- forwarded by mary hain / hou / ect on 03 / 14 / 2000 0	
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business breaking news alertdigisound inc . (nqb : dgsu)music rights verification , distribution a vince , stinson , looks like just me and jim irvine attending this one mays suffice since th david ,here are the reviews of the ut energy finance program that you requested :from alertfyi- - - - original message - - - -from : webmaster @ cera . com @ enronsent :	SPAM
----- forwarded by lorna brennan / et & s / enron on 12 / 14 executive summary? davis struggles to find senate support for socal transmission deal daily traders reportnxahxdbreaking newsvhcxwevlefsqwle - directto lower prices to re since lisa used fuchsia , i ' ll use lavender ..on # 1 i would eliminate that processing . what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho latest western markets forecast from cambridge energy .lorna- - - - original messa brand new acquisition , reverse merger , ground floor !start watching this company rig please share the following announcement with your associates .georgia institute of te to their first time / titlemeta http - equiv = content - type content = text / html ; chars this story was sent to you by peggy . mccurley @ enron . com with these comments :t mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i thi fyi .in the attached article , gov davis mentions that enron is one of the generators tha	SPAM
it was the spring of 1979 .i was just a tall , goofy looking kid in middle schoolwith buc cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we we prepared a presentaton on argentina and brazil for your sp trip . michael guerrero hello shmuエル ,thanks for your message . the end of 2000 and the beginning of 2001 w jeff ,we shall forward to you shortly a copy of the message from sandeep with thenun the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa mineral exploration stockpdelta mining and exploration corp . otc : dmxxppcontrols 6 p	SPAM

Message	Investments	Get Rich	As Seen As
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc 's proposed decision in the gas industry restructuring proce	SPAM	SPAM	
fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0			
- - - - - forwarded by mary hain / hou / ect on 03 / 14 / 2000 0			
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c	SPAM		
goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business	SPAM		
breaking news alertdigsound inc . (nqb : dgsu)music rights verification , distribution a		SPAM	
vince , stinson , looks like just me and jim irvine attending this one mays suffice since th			
david , here are the reviews of the ut energy finance program that you requested : from			
alertyi - - - - original message - - - - from : webmaster @ cera . com @ enronsent :			
- - - - - forwarded by lorna brennan / et & s / enron on 12 / 14	SPAM		
executive summary? davis struggles to find senate support for socal transmission dea			
daily traders reportnx fahxd breaking newsvhxcwevlefsqwe - directto lower prices to re			
since lisa used fuchsia , i ' ll use lavender . . on # 1 i would eliminate that processing .			
what is tmxt technologies ? tmxt issued 2 news releases today , one during market ho	SPAM	SPAM	
latest western markets forecast from cambridge energy . lorna- - - - original messa		SPAM	
brand new acquisition , reverse merger , ground floor ! start watching this company rig			
please share the following announcement with your associates . georgia institute of te			
to their first time / titlemeta http - equiv = content - type content = text / html ; chars			
this story was sent to you by peggy . mccurley @ enron . com with these comments : t			
mona vang , mineral exploration stockdelta mining and exploration corp . otc : dmxpco		SPAM	
thanks dr . kaminski , i forgot that : -) i ' ll understand it asi may identify work that i thi			
fyi . in the attached article , gov davis mentions that enron is one of the generators tha			
it was the spring of 1979 . i was just a tall , goofy looking kid in middle schoolwith buc			
cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we			
we prepared a presentaton on argentina and brazil for your sp trip . michael guerrero	SPAM		
hello shmuels , thanks for your message . the end of 2000 and the beginning of 2001 w			
jeff , we shall forward to you shortly a copy of the message from sandeep with thenun			
the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa			
mineral exploration stockdelta mining and exploration corp . otc : dmxpcontrols 6 p			

Message	Investments	Get Rich	As Seen As	MatchEmail
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring proce	SPAM	SPAM		
fyi .----- forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0			HAM	
----- forwarded by mary hain / hou / ect on 03 / 14 / 2000 0			HAM	
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c	SPAM		HAM	
goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business	SPAM		HAM	
breaking news alertdigisound inc . (nqb : dgsu)music rights verification , distribution a		SPAM	HAM	
vince , stinson , looks like just me and jim irvine attending this one mays suffice since th				
david , here are the reviews of the ut energy finance program that you requested :from				
alertyi ----- original message ----- from : webmaster @ cera . com @ enronsent :			HAM	
----- forwarded by lorna brennan / et & s / enron on 12 / 14		SPAM		
executive summary? davis struggles to find senate support for socal transmission dea			HAM	
daily traders reportnxfaahdbreaking newsvhxcwevlefsqwe - directto lower prices to re			HAM	
since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that processing .			HAM	
what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM	SPAM		
latest western markets forecast from cambridge energy .lorna----- original messa		SPAM		
brand new acquisition , reverse merger , ground floor !start watching this company rig			HAM	
please share the following announcement with your associates .georgia institute of te			HAM	
to their first time / titlemeta http - equiv = content - type content = text / html ; chars				
this story was sent to you by peggy . mccurley @ enron . com with these comments :t			HAM	
mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco		SPAM		
thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i thi				
fyi .in the attached article , gov davis mentions that enron is one of the generators tha			HAM	
it was the spring of 1979 .i was just a tall , goofy looking kid in middle schoolwith buc			HAM	
cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we			HAM	
we prepared a presentaton on argentina and brazil for your sp trip . michael guerriero	SPAM			
hello shmuels,thanks for your message . the end of 2000 and the beginning of 2001 w				
jeff ,we shall forward to you shortly a copy of the message from sandeep with thenun				
the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa			HAM	
mineral exploration stockdelta mining and exploration corp . otc : dmxpcontrols 6 p				

Message	Investments	Get Rich	As Seen As	MatchEmail	Votes
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring proce	SPAM	SPAM			SPAM
fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0				HAM	HAM
- - - - - forwarded by mary hain / hou / ect on 03 / 14 / 2000 0				HAM	HAM
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c	SPAM			HAM	Abstain
goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business	SPAM			HAM	Abstain
breaking news alertdigisound inc . (nqb : dgsu)music rights verification , distribution a		SPAM		HAM	Abstain
vince , stinson , looks like just me and jim irvine attending this one may suffice since th					Abstain
david , here are the reviews of the ut energy finance program that you requested :from					Abstain
alertfyi - - - - original message - - - - from : webmaster @ cera . com @ enronsent :				HAM	HAM
- - - - - forwarded by lorna brennan / et & s / enron on 12 / 14	SPAM				SPAM
executive summary? davis struggles to find senate support for socal transmission dea				HAM	HAM
daily traders reportnxahxdbreaking newsvhxcwevlefsqwe - directto lower prices to re				HAM	HAM
since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that processing .				HAM	HAM
what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM	SPAM			SPAM
latest western markets forecast from cambridge energy .lorna- - - - original messa	SPAM				SPAM
brand new acquisition , reverse merger , ground floor !start watching this company rig				HAM	HAM
please share the following announcement with your associates .georgia institute of te				HAM	HAM
to their first time / titlemeta http - equiv = content - type content = text / html ; chars					Abstain
this story was sent to you by peggy . mcurley @ enron . com with these comments :t				HAM	HAM
mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco		SPAM			SPAM
thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i thi					Abstain
fyi .in the attached article , gov davis mentions that enron is one of the generators tha				HAM	HAM
it was the spring of 1979 .i was just a tall , goofy looking kid in middle schoolwith buc				HAM	HAM
cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we				HAM	HAM
we prepared a presentaton on argentina and brazil for your sp trip . michael guerriero	SPAM				SPAM
hello shmuels,thanks for your message . the end of 2000 and the beginning of 2001 w					Abstain
jeff ,we shall forward to you shortly a copy of the message from sandeep with thenun					Abstain
the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa				HAM	HAM
mineral exploration stockpdelta mining and exploration corp . otc : dmxxpcontrols 6 p					Abstain

- Label Model
 - Simple Majority Vote
- Now have labels for **some** of the training data

Message	Votes
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring process . fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 00 : 00 : 00	SPAM
----- forwarded by mary hain / hou / ect on 03 / 14 / 2000 00 : 00 : 00	HAM
----- original message -----from : webmaster @ cera . com @ enronsent : 03 / 14 / 2000 00 : 00 : 00	HAM
----- forwarded by lorna brennan / et & s / enron on 12 / 14 / 2000 00 : 00 : 00	SPAM
executive summary? davis struggles to find senate support for socal transmission deal	HAM
daily traders report nxfahxdbreaking news vhcwevlefsqwl e - direct to lower prices to re	HAM
since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that processing .	HAM
what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM
latest western markets forecast from cambridge energy .lorna- - - - original messa	SPAM
brand new acquisition , reverse merger , ground floor !start watching this company rig	HAM
please share the following announcement with your associates .georgia institute of te	HAM
this story was sent to you by peggy . mccurley @ enron . com with these comments : t	HAM
mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco	SPAM
fyi .in the attached article , gov davis mentions that enron is one of the generators tha	HAM

Some Labeled data!

	Votes
be worth \$ 50 , 000 in a short period of time . read more	SPAM
s proposed decision in the gas industry restructuring proce	HAM
thuraisingham / enron communications on 02 / 29 / 00 C	HAM
-- forwarded by mary hain / hou / ect on 03 / 14 / 2000 C	HAM
age -----from : webmaster @ cera . com @ enronsent :	HAM
-- forwarded by lorna brennan / et & s / enron on 12 / 14	SPAM
truggles to find senate support for socal transmission deal	HAM
breaking newsvhxcwevlefsqwle - directto lower prices to re	HAM
use lavender . .on # 1 i would eliminate that processing .	HAM
mxt issued 2 news releases today , one during market ho	SPAM
caset from cambridge energy .lorna- - - - original messa	SPAM
rse merger , ground floor !start watching this company rig	HAM
nouncement with your associates .georgia institute of te	HAM
peggy . mccurley @ enron . com with these comments :t	HAM
ction stockdelta mining and exploration corp . otc : dmxpco	SPAM
gov davis mentions that enron is one of the generators tha	HAM



Classification
Algorithm (Logistic
Regression . . .
BERT)

		Predicted	
		HAM	SPAM
Actual	HAM	45	5
	SPAM	14	6

F1: 83%

The Code is never done

- Misspellings
- Known
Safelist/Good list of
addresses

```
if Misspell_Model=True:  
    label(SPAM)
```

```
if Dict_Good=True:  
    label(HAM)
```

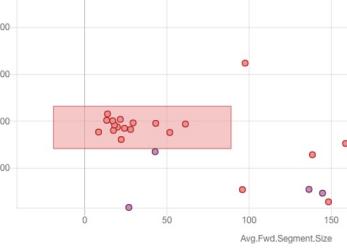


Concepts

Formula / Code

Many types of functions you can use

	A	B
1	Pattern	
2	\b[A-Z]{2}\d{3}\b	
3		
4	String	
5	Chain frame CH-110 I/O 1034567 \$10.99	SKU available
6	20x Leaf puzzle LF-020 I/O 1034567 \$7.49	TRUE
7	Metal shelf \$49.99 I/O 1033129 23-MAR-2022	FALSE
8	GL 100 Golder Units, I/O 1034567, \$25.50	FALSE
9	Case sandpaper, I/O 1033877, CS-190, \$15.00	TRUE



Text Patterns

Visual/Numeric Patterns



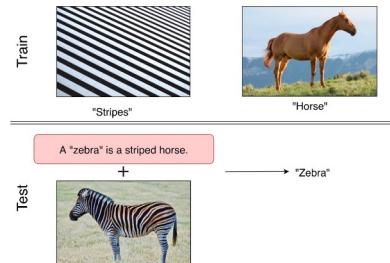
Dictionaries



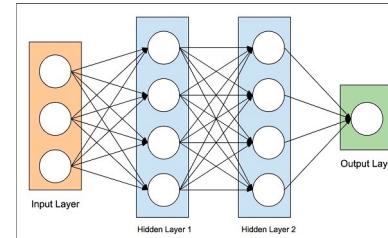
Database Lookups



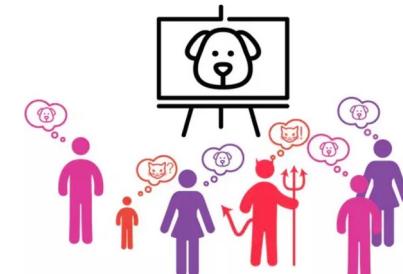
Automated LFs



Zero Shot Learning



External Models



Crowdsource

Message	Investments	Get Rich	As Seen As	MatchEmail	Misspell_Model
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring proce	SPAM	SPAM		HAM	SPAM
fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0				HAM	
- - - - - forwarded by mary hain / hou / ect on 03 / 14 / 2000 0				HAM	SPAM
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c	SPAM			HAM	
goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business	SPAM			HAM	
breaking news alertdigisound inc . (nqb : dgsu)music rights verification , distribution a			SPAM	HAM	
vince , stinson , looks like just me and jim irvine attending this one mays suffice since th					
david , here are the reviews of the ut energy finance program that you requested : from					SPAM
alertfyi- - - - original message -----from : webmaster @ cera . com @ enronsent :				HAM	
- - - - - forwarded by lorna brennan / et & s / enron on 12 / 14		SPAM			SPAM
executive summary? davis struggles to find senate support for socal transmission dea				HAM	
daily traders reportnxfahxdbreaking newsvhxcwevlefsqwl - directto lower prices to re				HAM	
since lisa used fuchsia , i ' ll use lavender .. on # 1 i would eliminate that processing .				HAM	SPAM
what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM	SPAM			SPAM
latest western markets forecast from cambridge energy .lorna- - - - original messa		SPAM			SPAM
brand new acquisition , reverse merger , ground floor !start watching this company rig				HAM	
please share the following announcement with your associates .georgia institute of te				HAM	
to their first time / titlemeta http - equiv = content - type content = text / html ; chars					
this story was sent to you by peggy . mccurley @ enron . com with these comments : t				HAM	
mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco			SPAM		SPAM
thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i thi					SPAM
fyi .in the attached article , gov davis mentions that enron is one of the generators tha				HAM	
it was the spring of 1979 .i was just a tall , goofy looking kid in middle schoolwith buc				HAM	
cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we				HAM	SPAM
we prepared a presentaton on argentina and brazil for your sp trip . michael guerrero	SPAM				SPAM
hello shmuels,thanks for your message . the end of 2000 and the beginning of 2001 w					
jeff ,we shall forward to you shortly a copy of the message from sandeep with thenun					
the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa				HAM	
mineral exploration stockdelta mining and exploration corp . otc : dmppcontrols 6 p					SPAM

Message	Investments	Get Rich	As Seen As	MatchEmail	Misspell_Model	Dict_Good
a \$ 3 , 000 investment could be worth \$ 50 , 000 in a short period of time . read more attached below is the cpuc ' s proposed decision in the gas industry restructuring proce	SPAM	SPAM		HAM	SPAM	
fyi . - - - - forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 0				HAM		HAM
- - - - - forwarded by mary hain / hou / ect on 03 / 14 / 2000 0				HAM	SPAM	HAM
enjoy media (ejym) , a chinese media company , on fast growth track by signing up c	SPAM			HAM		HAM
goodday , i am frenkel heydorn , staff of smith came to our bank to engage in business	SPAM			HAM		HAM
breaking news alertdigisound inc . (nqb : dgsu)music rights verification , distribution a			SPAM	HAM		HAM
vince , stinson , looks like just me and jim irvine attending this one mays suffice since th						
david , here are the reviews of the ut energy finance program that you requested :from					SPAM	
alertyfyi - - - original message - - - from : webmaster @ cera . com @ enronsent :				HAM		
- - - - - forwarded by lorna brennan / et & s / enron on 12 / 14		SPAM			SPAM	
executive summary? davis struggles to find senate support for socal transmission dea				HAM		HAM
daily traders reportnx fahxd breaking newsvhxcwevlefsqwe - directto lower prices to re				HAM		HAM
since lisa used fuchsia , i ' ll use lavender . .on # 1 i would eliminate that processing .				HAM	SPAM	
what is tmxt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM	SPAM			SPAM	
latest western markets forecast from cambridge energy .lorna- - - original messa		SPAM			SPAM	
brand new acquisition , reverse merger , ground floor !start watching this company rig				HAM		
please share the following announcement with your associates .georgia institute of te				HAM		HAM
to their first time / titlemeta http - equiv = content - type content = text / html ; chars						
this story was sent to you by peggy . mccurley @ enron . com with these comments :t				HAM		
mona vang ,mineral exploration stockdelta mining and exploration corp . otc : dmxpco			SPAM		SPAM	
thanks dr . kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i thi					SPAM	
fyi .in the attached article , gov davis mentions that enron is one of the generators tha				HAM		HAM
it was the spring of 1979 .i was just a tall , goofy looking kid in middle schoolwith buc				HAM		HAM
cera alert : december 13 , 2000title : california on the brinkcera knowledge areas : we				HAM	SPAM	
we prepared a presentaton on argentina and brazil for your sp trip . michael guerrero	SPAM				SPAM	
hello shmuels,thanks for your message . the end of 2000 and the beginning of 2001 w						
jeff ,we shall forward to you shortly a copy of the message from sandeep with thenun						
the stock to buylearn what jp morgan , sg cowan and the wall street journal had to sa				HAM		
mineral exploration stockdelta mining and exploration corp . otc : dmxpcontrols 6 p					SPAM	HAM

	Investments	Get Rich	As Seen As	MatchEmail	Misspell_Model	Dict_Good	Votes
orth \$ 50 , 000 in a short period of time . read more	SPAM	SPAM		HAM	SPAM		SPAM
osed decision in the gas industryrestructuring proce				HAM		HAM	HAM
isingham / enron communications on 02 / 29 / 00 0				HAM		HAM	HAM
warded by mary hain / hou / ect on 03 / 14 / 2000 0				HAM	SPAM	HAM	HAM
edia company , on fast growth track by signing up c	SPAM			HAM		HAM	HAM
aff of smith came to our bank to engage in business	SPAM			HAM		HAM	HAM
(nqb : dgsu)music rights verification , distribution a		SPAM		HAM		HAM	HAM
and jim irvine attending this one may suffice since th							Abstain
ut energy finance program that you requested :from					SPAM		SPAM
-- -from : webmaster @ cera . com @ enronsent :				HAM			HAM
warded by lorna brennan / et & s / enron on 12 / 14	SPAM				SPAM		SPAM
es to find senate support for socal transmission dea				HAM		HAM	HAM
g newsvhxcwevlefsqwl - directto lower prices to re				HAM		HAM	HAM
vender .on # 1 i would eliminate that processing .				HAM	SPAM		Abstain
sued 2 news releases today , one during market ho	SPAM	SPAM			SPAM		SPAM
from cambridge energy .lorna - - - original messa		SPAM			SPAM		SPAM
erger , ground floor !start watching this company rig				HAM			HAM
cement with your associates .georgia institute of te				HAM		HAM	HAM
- equiv = content - type content = text / html ; chars							Abstain
y . mccurley @ enron . com with these comments :t				HAM			HAM
ockdelta mining and exploration corp . otc : dmxpco		SPAM			SPAM		SPAM
-) i ' ll understand it asi may identify work that i thi					SPAM		SPAM
vis mentions that enron is one of the generators tha				HAM		HAM	HAM
st a tall , goofy looking kid in middle schoolwith buc				HAM		HAM	HAM
e : california on the brinkcera knowledge areas : we				HAM	SPAM		Abstain
entina and brazil for your sp trip . michael guerriero	SPAM				SPAM		SPAM
age . the end of 2000 and the beginning of 2001 w							Abstain
ly a copy of the message from sandeep with thenun							Abstain
gan , sg cowan and the wall street journal had to sa				HAM			HAM
ining and exploration corp . otc : dmxpcontrols 6 p					SPAM	HAM	Abstain

- As your functions increase
 - Overlaps
 - Conflicts
- Majority vote isn't optimal, can be other label model strategies

	Votes
investment could be worth \$ 50 , 000 in a short period of time . read more	SPAM
below is the cpuc ' s proposed decision in the gas industry restructuring proce	HAM
forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 C	HAM
----- forwarded by mary hain / hou / ect on 03 / 14 / 2000 (HAM
ia (ejym) , a chinese media company , on fast growth track by signing up c	HAM
am frenkel heydorn , staff of smith came to our bank to engage in business	HAM
ews alertdigsound inc . (nqb : dgsu)music rights verification , distribution :	HAM
e are the reviews of the ut energy finance program that you requested :from	SPAM
-- original message -----from : webmaster @ cera . com @ enronsent :	HAM
----- forwarded by lorna brennan / et & s / enron on 12 / 14	SPAM
summary? davis struggles to find senate support for so cal transmission deal	HAM
rs reportnxfahxdbreaking newsvhxcwevlefsqwle - directto lower prices to ru	HAM
xt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM
tern markets forecast from cambridge energy .lorna----- original messa	SPAM
acquisition , reverse merger , ground floor !start watching this company rig	HAM
re the following announcement with your associates .georgia institute of te	HAM
was sent to you by peggy . mccurley @ enron . com with these comments :t	HAM
,mineral exploration stockdelta mining and exploration corp . otc : dmxpco	SPAM
kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i th	SPAM
attached article , gov davis mentions that enron is one of the generators tha	HAM



Classification
Algorithm (Logistic
Regression . . .
BERT)

	Votes
investment could be worth \$ 50 , 000 in a short period of time . read more	SPAM
below is the cpuc ' s proposed decision in the gas industry restructuring proce	HAM
forwarded by ravi thuraisingham / enron communications on 02 / 29 / 00 C	HAM
----- forwarded by mary hain / hou / ect on 03 / 14 / 2000 (HAM
ia (ejym) , a chinese media company , on fast growth track by signing up c	HAM
am frenkel heydorn , staff of smith came to our bank to engage in business	HAM
ews alertdigisound inc . (nqb : dgsu)music rights verification , distribution :	HAM
e are the reviews of the ut energy finance program that you requested :from	SPAM
-- original message -----from : webmaster @ cera . com @ enronsent :	HAM
----- forwarded by lorna brennan / et & s / enron on 12 / 14	SPAM
summary? davis struggles to find senate support for socal transmission deal	HAM
rs reportnxfahxdbreaking newsvhxcwevlefsqwl - directto lower prices to ru	HAM
xt technologies ?tmxt issued 2 news releases today , one during market ho	SPAM
tern markets forecast from cambridge energy .lorna----- original messa	SPAM
acquisition , reverse merger , ground floor !start watching this company rig	HAM
re the following announcement with your associates .georgia institute of te	HAM
was sent to you by peggy . mccurley @ enron . com with these comments :t	HAM
,mineral exploration stockdelta mining and exploration corp . otc : dmxpco	SPAM
kaminski ,i forgot that : -)i ' ll understand it asi may identify work that i th	SPAM
attached article , gov davis mentions that enron is one of the generators tha	HAM



Classification
Algorithm (Logistic
Regression . . .
BERT)

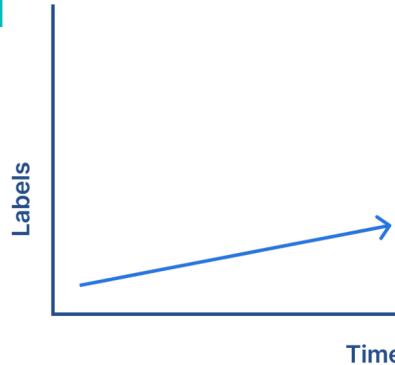
		Predicted	
		HAM	SPAM
Actual	HAM	46	4
	SPAM	6	14

F1: 90%

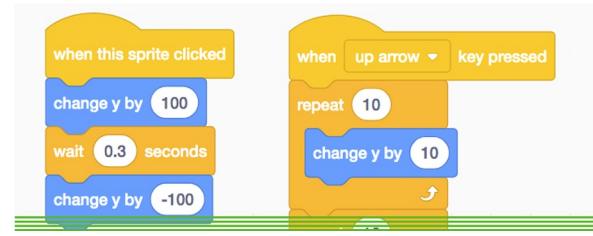
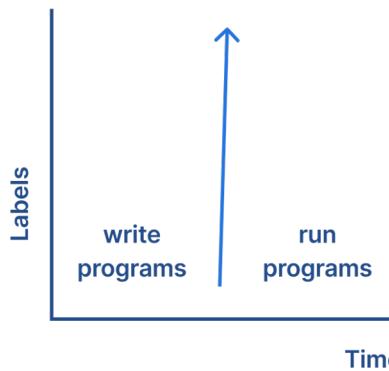
How this changes AI Development in Enterprises

Comparing Manual and Programmatic Labeling

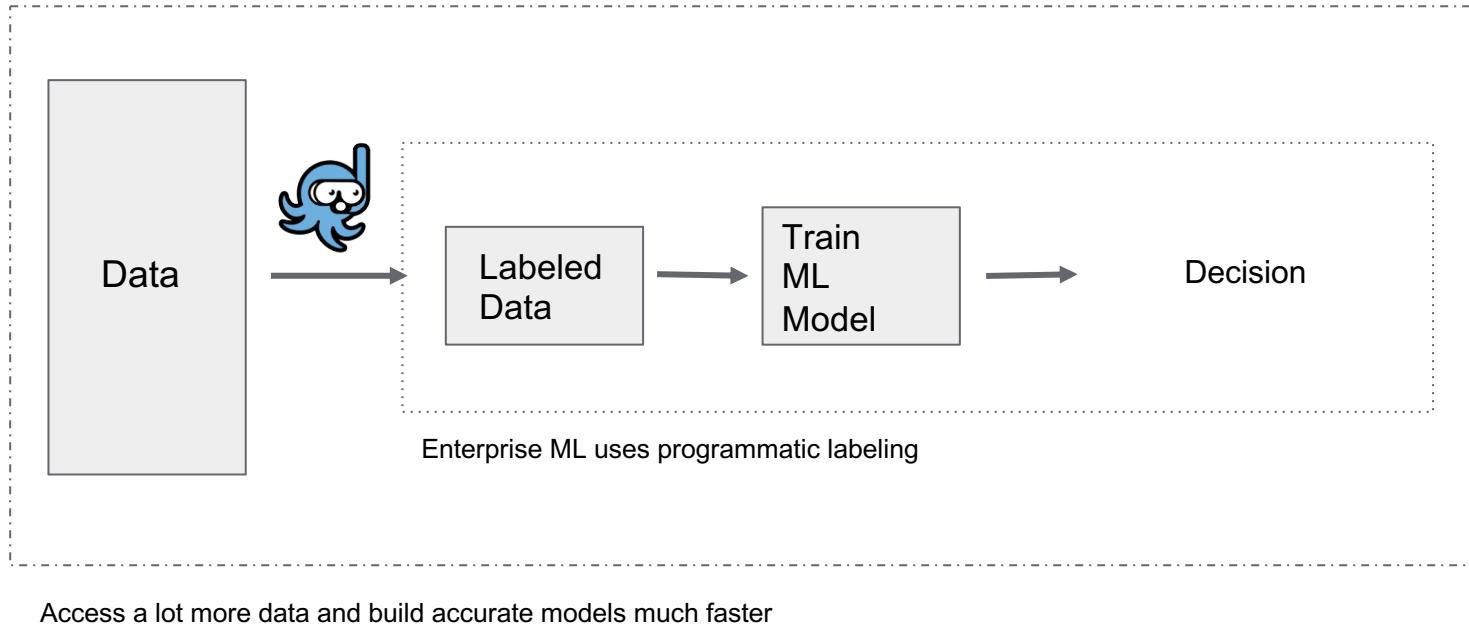
Manual



Programmatic



Snorkel Can Increase AI Development by 10x to 100X





Google needed topic and product classifiers



Sunglasses



Watch



Watch



Not Accessory



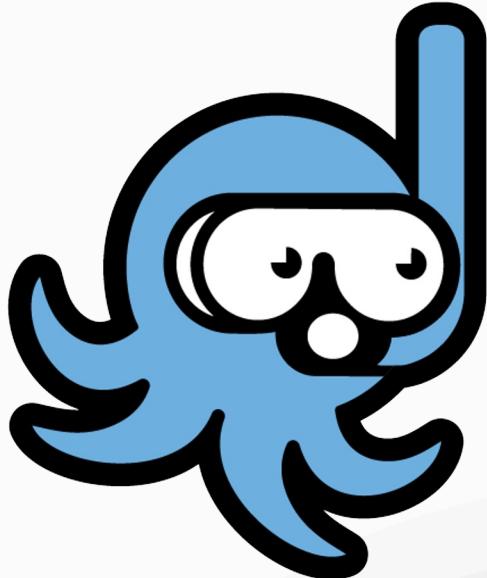
Not Accessory
Bag



Sunglasses

Google

Content Classification



Hours

Time to add a new class and re-label all data, from days

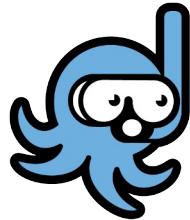
6.5MM

Data points labeled programmatically with Snorkel

52%

Average model performance improvement

The Snorkel Project



Google



Infosys

facebook

LinkedIn

BASF
We create chemistry



Microsoft

Chegg®

teradata.

accenture



TOSHIBA

vmware®



National Institutes of Health
Your Health
Our Discovery
Your Health



U.S. Department of Veterans Affairs



NEC

HITACHI



SAP



GORDON AND BETTY
MOORE
FOUNDATION

5+ year research project at the Stanford AI Lab resulting in
50+ publications and many production deployments

<https://github.com/snorkel-team/snorkel>

Next Steps

- Alternative to manual labeling
- Check out snorkel.ai website to learn more
- Go try programmatic labeling with Snorkel

Snorkel

- Rajiv Shah
- raj@snorkel.ai
- Linkedin, Twitter - @rajcs4, Tik Tok - @rajistics