M_	T	w	T	F	s	\$
Page	No.:				.	UVA
Date:						

-

M T W T F S S Page No.: YOUVA
Date:
Maine Bayes on Text Data
- Performs super good on textual data as
per deserved
The contract of the contract o
Ex ue have data about revieue and
sentiment
The second of th
renieur sontiment
- Carrear
and the second of the second o
on a strong of the same of the
Step ! Text Proprocessing
in the state of th
such as -> Remove html tags
-) " special characters.
-) Converting everything to lower case
Removino Stephiono
- Stemming
(x) the his production of the property
Step 2: Ve ctorizing (text to numbers)
Contraction of the second of t
· Different decomiques to lectorizitation
- BOW
+ ONE The problem of the problem of the second
TEID was the second of the second
Em bodding " Wille in in.
The first of the second of the
Step 3: apply Main Bayer

м	T	w	T	F	s	s
Page No.:						
Date:					1	AVU
Date:						

Bag	0	1. 4/09	1ds (Lectoriz	ation)
7		A		,	

Eλ	reviews	Sinamore	
	liked the movie	tre	
ie v	hated the movil	- Ug	
	7. 2	13155	the state of the s
	_		

· BOW will include all the available

· Let's say me have lo lath monds

You see in that 10 lake there were 35000 unique words

· Ond in among 35000, let's say you will choose 5000 most frequent words

Ex: hote like, mouse actor. - 5000 word

review and ask how many times
a genieum contain hate, like, and all that
5000 words

Cx	Most frequent word	s thate	elite	mouie	1. 5120,642 1
	Review 1	2 0	()	13	
	Rujew 2	2	0	3	J
	Review 3		3	4 4	
			To be	Account.	`
				1	No.
	Review 50000)		

		M T W T Page No.: Date:	F S S YOUVA
· Noue we have the a	ataset, wi	hich con	tains
50000 nows and	5000 colu	mm.	
(Vectorization)			
• Atala O II	Laine and	- 1/4 1/4	94.
· Nobe let's say un	nau a g	rew que	ng 💗
(surfice), un ham	to pea		*
Sentiment			•
· Review was " Adipunu	sh is quat	mouil	
	it way v		•
* F ~ ~	0		
· Noue we will tou	to supress	ent tex	× .
in sevieu some		olum ns	W •
have 's	rat contra	, , , , , , , , , , , , , , , , , , , ,	•
· Me had 5000 colums	ns '.		•
Ruieu hate like movie	00109	r l	•
dipurush: 0 2 3	actol		•
		1	
· This tells how many	times word	hate, /	ik
This tells how many actor has come in	service of	adirector	sh
Digital to the state of	130 / 100	2. 11.	•
· Basically, un convent	ed into	Wector.	•
· Basically un convent of dimension -	(1,5000)		•
V			•
· Nous us have to	tell 16.	thi.	•
grusy is ptus 08	-uo	111.	•
• III thank de la lite	0	· Control	
· we have , to find	· 2 propa	biletis.	
	/	,	
8 Cu	- he	in and	
		21 31.	

e No.:	AVUOY	
e :	10072	
=0, liki=2.)
	,	

MTWTFSS

P(+ve | hate = 0, like = 2.....) & P(-ve | hate

· similarly we will calculate the - w prob

P (tre | hade = 0, like = 2) = P (hade = 0 | tre) X P (11ke = 2 | tre) -- : XP (tre)

· let's say we have the prob = 0.8,

honce use will adipurush santera positive

Data - 22-Mad