A Charles of the self	
	Assignment 4
-	Title: Stemming and pealure selection: techniques using vectors
-	uchniques using vectors
	Peroblem statement: consider a enitable tent.
	anny semmina
	demine selection techniques to represent
9	downents as vectors, classify documents as nectors precision and recall.
The Real Property lies	
-	Objective: 1) 3 implementation of peroblem using
The state of the s	python .
	2) Remove stop words, applying stemming and feature selection.
	general succession.
And the state of	· Outcome: i) Understanding the stemming
13 1	and feature selection process. 2) learn about precision and execul.
0	2) learn about precision and execall.
	Theory:
	2111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0/12	1) stop words:
	In computing stop mords are mords
1990000	processing of natural language data (text). Though stop words are usually reports
	processing of natival language sata cient.
	the most common values or words in
	a language, there is no universal
	a language, there is no universal list of stop words used by all

natural processing tools. Example: for search engines: the, is, at, which, on, etc 2) Stemming: Stemming is the process of reducing inflected The stem need not be identical to the mountelogical wood of the word Many search engines treat words with some sterns as synonyms as a kind of query expansion, a process called: conflation Feature Selection: In machine learning and statistics feature selection also known as nariable subset selection, attribute ou variable subset selection relevant features (unriables, prudictions) for use in model construction

Aschitecture Feature Selection post tagged sentences or evitaly naw Text string Sentence Segmentation relation (tokenization leasons por use of feature selection technique. D simplification of models to make easy to interpret 2) lesser training time 3) To avoid curse of dimensionality 4) Enhanced generalization by reducing overfitting

	V	4 1		
	Test Cases:	Salalak sa	Table of a property	
	special publication			
See	Descript	ton =	Expected	Actual
no	A smerty fire		0/P	0/8
1	Import pane		337797712	0
	other liber		Success	Success
	jupytese no	lebook	73313357352	0
2	Preprocess	data	success	euces
	raprocess		milway state	
3	Apply suffi	x strimming		
	Apply suffi	algorithm	success	succes
	Janes	J	F-11107-8	
4	Divide da	ta into train	As Mart	
		and find	Month success	Succes
	accuracy			
			Tay / moons?	
Co	nclusion: Th	us we ha	ne studied to	remove
٨	top words,	apply sle	morning and expressed doci	Bealtice
	election techn	igues to	represent don	mens
	as vectors.		Ministry Portey of	
				7
			Land To V	