1) Choose all the correct statements about the dataset at the following URL 1 pr	pint
https://huggingface.co/datasets/ai4bharat/naamapadam 1	
Note: We strongly recommend using the appropriate functions in the disaset, module to obtain the dataset's information, rather than accessing that information directly from the hub.	
Note: We strongly recommend using the appropriate functions in the dataset module to obtain the dataset's information, rather than accessing that information directly from the hub.	
The dataset contains 11 configs(subsets)	
The dataset contains 22 configs(subsets)	
☐ The dataset contains 22 contigs(subsets)	
The Hindi subset of the dataset contains 985787 training samples	
The Tamil subset of the dataset contains a total of 985787 samples.	
The number of classes (labels,tags) is 7	
2) Common data	point
Provided by Tarillands days of females (ID) have all the single and (Alberta Community and the single and the	
Download the Tamil sub-dataset from the following URL https://huggingface.co/datasets/ai4bharat/naamapadam, and store in a variable named "ds".	
Find the location of the cache directory (one way to determine the cache directory location is by using a method from the Dataset class; please refer to the documentation). Which of the following files are in the cache	d
directory corresponding to the downloaded dataset?	
naamapadam-train.arrow since there are 3 types of setting in the	
naamapadam-test.arrow Since there are 3 types of setting in the dataset -> terain test realidad	
naamapadam-validation.arrow	
but also if you that the forum a jour	
— naamapadam-all.arrow	الم
file is created that stones the info. sugar. the dataset.	~~~~
the latant	U
South Market .	
3) What is the disk space size of the downloaded dataset in MB?	
180	
	<u> </u>
Yes, the answer is correct.	
Score: 1	
Accepted Answers:	
(Type: Range) 179,181	point
4) Enter the number of training samples in the dataset. \longrightarrow druck NB	
-	
497882	
47/002	
17/102	
	do"
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable.	ds".
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable "How many tokens (in millions) are there in the entire dataset?	ds*.
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable "How many tokens (in millions) are there in the entire dataset?	ds*.
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable "How many tokens (in millions) are there in the entire dataset?	ds*.
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable "How many tokens (in millions) are there in the entire dataset?	ds".
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable "How many tokens (in millions) are there in the entire dataset?	ds".
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	ds".
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	ds*.
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	ds".
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alaund 6 Millions	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alabard 6 millions 7 Indicate the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alabard 6 millions 6) The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is column.	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Abound 6 millions 7 Interpretation of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Score: 0 Accepted Answers:	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alabard 6 millions 7 Indicate the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Alabard 6 millions 6) The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is column.	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Abound 6 millions 7 Interpretation of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Score: 0 Accepted Answers:	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the efficie dataset? 4 5 All aund 6 millions 6) The statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of The statement that the modified dataset increased the disk space requirement about two fold is usuards. 8 False No, the answer is incorrect. Score: 0 Accepted Answers: True	
5) Create a new column named "num tokens" Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Abraund 6 millions 7 Interpretation of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Calumn with the number of tokens in the statement that the modified dataset to the same variable? 7) Concatenate all the samples across the splits in the following order: [traintest validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the efficie dataset? 4 5 All aund 6 millions 6) The statement that the modified dataset increased the disk space requirement about two fold is Calumn with total no. of The statement that the modified dataset increased the disk space requirement about two fold is usuards. 8 False No, the answer is incorrect. Score: 0 Accepted Answers: True	
5) Create a new column named 'num tokens' Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable 'How many tokens (in millions) are there in the entire dataset? 4 5 6 Abound 6 millions 7 7 8 The district dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space as a statement in the same with the modified dataset increased the disk space as a delimiter and store the resulting sample on the same variable 'ds'. Eiter the total number of samples in 'ds'. Store the new dataset in the same variable 'ds'. Eiter the total number of samples in 'ds'.	
5) Create a new column named 'num tokens' Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable' How many tokens (in millions) are there in the entire dataset? 4 5 6 7 6 The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Words False No, the answer is incorrect. Score: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (folores) in each sample using a single white space as a delimiter and store the resulting sample in a new column named 'text'. Create a new dataset by removing the columns 'ner tage' and 'Tokens'. Store the new dataset in the same variable 'ds'. Enter the total number of samples in 'ds'. 501435	
5) Create a new column named 'num tokens'. Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable' how many tokens (in millions) are there in the entire dataset? 4 5 Accounted 6 millions 7 Madurud 6 millions 6) The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of The Statement that the modified dataset increased the disk space requirement about two fold is Column with total number of the same variable's not sample using a single white space as a delimiter and store the resulting sample in a new column named 'text'. Create a new dataset by removing the columns 'ner tage' and 'tokens', Store the new dataset in the same variable's'. Store the new dataset in the same variable's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the columns 'ner tage' and 'tokens', Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable 's'. Store the new dataset in the same variable	
5) Create a new column named 'num tokens' Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable' How many tokens (in millions) are there in the entire dataset? 4 5 6 7 6 The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Words False No, the answer is incorrect. Score: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (folores) in each sample using a single white space as a delimiter and store the resulting sample in a new column named 'text'. Create a new dataset by removing the columns 'ner tage' and 'Tokens'. Store the new dataset in the same variable 'ds'. Enter the total number of samples in 'ds'. 501435	
5) Create a new column named "num tokens" (ompute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest/validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens.) Store the new dataset in the same variable of the same variable of the same variable of the same variable. Store: 1 7) Concatenate all the samples across the splits in the following order: [traintest/validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens.) Store the new dataset in the same variable of the sa	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the active created column. Reassing the modified dataset to the same variable. How many tokens (in millions) are there in the emitre dataset? 4 5 Accepted Answers: True To Concatenate all the samples across the splits in the following order: [train-test-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (fokens) in each sample using a single white space as a delimiter and store the resulting sample in a new column named "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new dataset in the same variable "test". Create a new dataset by removing the columns "ner tags" and "tokens". Store the new data	
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the anewly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the millions) are there in the millions are the mi	int
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the active created column. Reassing the modified dataset to the same variable. How many tokens (in millions) are there in the emitre dataset? 4 5 Abound 6 millions 7 The statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two fold is Column with total no. of the statement that the modified dataset increased the disk space requirement about two folds to column named. The statement that the modified dataset increased the disk space requirement about two folds to column named. The statement that the modified dataset increased the disk space requirement about two folds. 4 5 The statement that the modified dataset increased the disk space requirement about two folds. The statement that the modified dataset increased the disk space requirement about two folds. 5 The statement that the modified dataset increased the disk space requirement about two folds. 6 The statement that the modified dataset increased the disk space requirement about two folds. 7 Concatenate all the samples across the splits in the following order: [train-test-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each sample sample and store the resulting sample in a new column named "test". Create a new dataset by removing the columns "rer tags" and "tokens" Store the new dataset in the same variable. The same variable of the same variable of	int
5) Create a new column named "num tokens" Compute the number of tokens (words) in each sample and store the results in the newly created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the entire dataset? 4 5 6 7 7 7 7 7 7 7 7 7 7 7 7	plant
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the new created column. Reassign the modified dataset to the same variable. How many tokens (in millions) are there in the millions) are there in the millions are there in the many tokens (in millions). 7	plant
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the accepted column. Reason the modified dataset to the same variable. 4 5 Al a und 6 "I would be not been accepted a fast of the statement that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Sore: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each sample using a single white space as a delimiter and store the resulting sample in a new column named "text". Create a new dataset by removing the columns "for tags" and "tokens". Store the new dataset in the same variable is: Yes, the answer is correct. Socre: 1 Accepted Answers: (Type Nameric) 501435	plant
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the accepted column. Reason the modified dataset to the same variable. 4 5 Al a und 6 "I would be not been accepted a fast of the statement that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Sore: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each sample using a single white space as a delimiter and store the resulting sample in a new column named "text". Create a new dataset by removing the columns "for tags" and "tokens". Store the new dataset in the same variable is: Yes, the answer is correct. Socre: 1 Accepted Answers: (Type Nameric) 501435	plant
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the accepted column. Reason the modified dataset to the same variable. 4 5 Al a und 6 "I would be not been accepted a fast of the statement that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Sore: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each sample using a single white space as a delimiter and store the resulting sample in a new column named "text". Create a new dataset by removing the columns "for tags" and "tokens". Store the new dataset in the same variable is: Yes, the answer is correct. Socre: 1 Accepted Answers: (Type Nameric) 501435	plant
5) Create a new column named "num tokens". Compute the number of tokens (words) in each sample and store the results in the accepted column. Reason the modified dataset to the same variable. 4 5 Al a und 6 "I would be not been accepted a fast of the statement that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is Column with that the modified dataset increased the disk space requirement about two fold is False No, the answer is incorrect. Sore: 0 Accepted Answers: True 7) Concatenate all the samples across the splits in the following order: [traintest-validation]. Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual words (tokens) in each sample using a single white space as a delimiter and store the resulting sample in a new column named "text". Create a new dataset by removing the columns "for tags" and "tokens". Store the new dataset in the same variable is: Yes, the answer is correct. Socre: 1 Accepted Answers: (Type Nameric) 501435	plant
5) Create a new column named "rum tokens". Compute the number of tokens (words) in each sample and store the results in the servely created column. Reason the modified dataset to the same variable from many tokens (in millions) are there in the Brine dataset? 4 5 Accorded Answers: True 7) Concatenate all the samples across the optics in the following order: (trainstervallidation). Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual action (tokens) in each sample sample sings a single white space as a delimiter and direct the resulting sample in a new column named "text". Create a line dataset by removing the columns "the tay" and "tokens". The value of the samples across the optics in the following order: (trainstervallidation). Currently, each sample contains a list of words (tokens). Create a sentence by joining the individual action (tokens) in each samples sample sample across the optics in the resulting sample in a new column named "text". Create a line of dataset by removing the columns "the tay" and "tokens". They the removing the columns "the tay" and "tokens". They the removing the columns "the tay" and "tokens". They the removing the columns "the tay" and "tokens". They the removing the columns "the tay" and "tokens". They they are samples across the spile as the samples across the spile across th	plant
5) Create a new column named "num tokens". Compute the number of tokens (vagreds) in each sample and store the results in the newly created column. Reason the modified dataset to the name variable "they many tokens" (in millions) are there in the Britis dataset? 4 4 5 Advanced 6 millions 7	plant

exact number. If	ses so that all the samples in the dataset should have at least six tokens (any symbol separated by a white space is considered a token). Enter the number of samples in the dataset after filtering. Enter the the answer is 123456, enter it as 123456.	7	
	the answer is 12345b, enter it as 12345b.	1	
370495			
Yes, the answe	er is correct.		
Score: 1		1	A 1~
Accepted Ans	wers:	- 1	Notebook
(Type: Numeric)	370495 1 point		Pusblem.
10) Download al	I the splits of Tamil sub-dataset "inltkh.ta" of https://huggingface.co/datasets/ai4bharat/indic_glue .	- 1	•
naamapadam ar	so that each sample contains at least six words (separated by a single white space). Then Interleave the resultant dataset with the above filtered dataset of naamapadam. Take 80% of samples from do use the contains at least six words (separated by a single white space). Then Interleave the resultant dataset with the above filtered dataset of naamapadam. Take 80% of samples from do use of the seed argument to 42.	J	
32354	-		