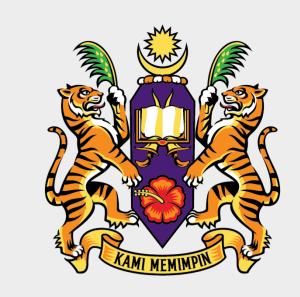


# Text Analytics of Course Reviews on Coursera Platform

Chan Huan Yang, Ramindhran Raja Mohan

School of Computer Science, University Science of Malaysia, Penang, Malaysia



spend lot time

### Introduction

We are now entering a new era - the revolution of online learning. From working professionals to recent high school graduates, many of them have found the reasons to take all or some of their courses online in platform such as Coursera, Udemy, and Edx









### **Problem Statement**

Ratings and reviews are always the major consideration factor by online course seekers before they joining the course. However, it can be time-consuming to read all the information especially the course reviews.

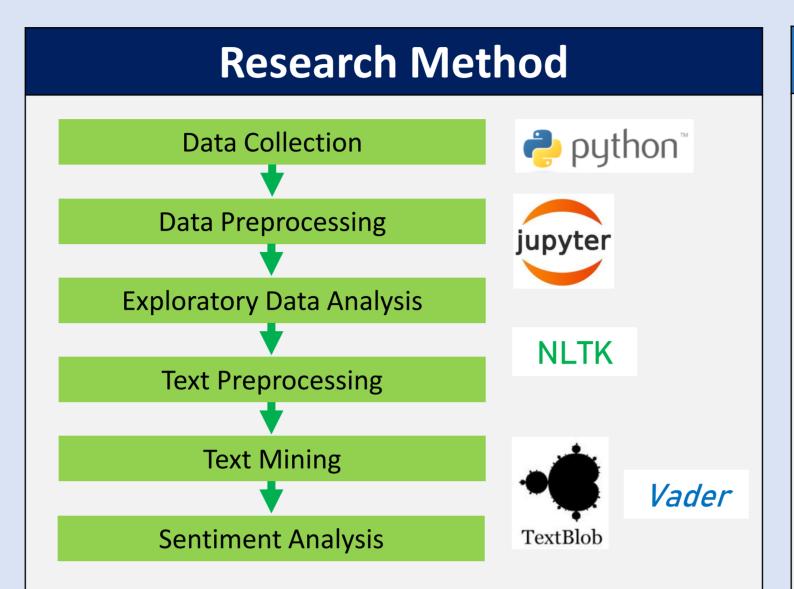
### **Research Question**

The research questions for this work were:

- How text analytics techniques such as ngram analysis, word cloud, and sentiment analysis can be applied to improve the online course searching process?
- What insights can be obtained by using text analytics techniques such as n-gram analysis, word cloud, and sentiment analysis?

# **Purpose of The Study**

Our objective is to propose a text analytics pipeline that includes text cleaning, text lemmatization, sentiment analysis, mining, and visualization that can help course seekers to gain a quick insight into the courses as well as enables them to make a quick comparison between multiple courses



### **Data Collection**

The data used in this work is from Kaggle.

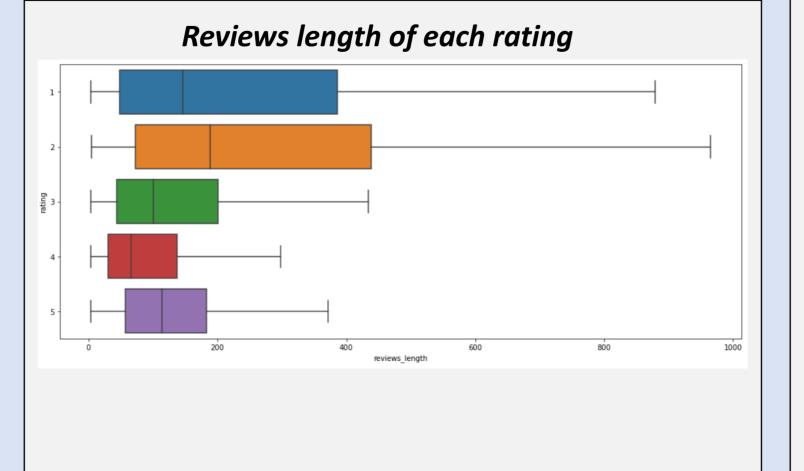
Course Name	URL	No of Review	
Programming for Everybody	https://www.coursera.org/l	45218	
(Getting Started with Python)	earn/python		
Python Data Structures	https://www.coursera.org/l	33543	
	earn/python-data		
Introduction to Data Science	https://www.coursera.org/l	14289	
in Python	earn/python-data-analysis		

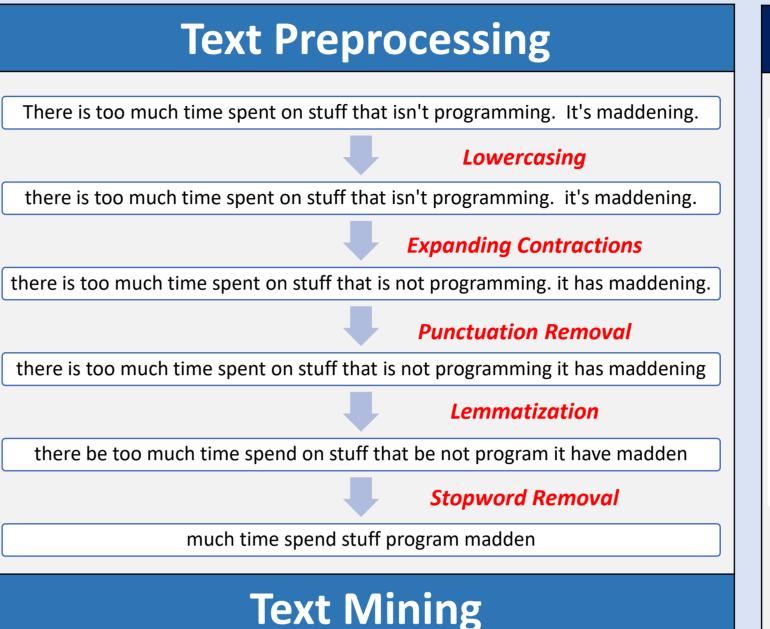
# **Data Preprocessing**

Data preprocessing steps include:

- removed the duplicate reviews
- removed the reviews with string's length less than three
- selected the English labeled reviews only using package "langid"

# **Exploratory Data Analysis**





### Word cloud of overall reviews

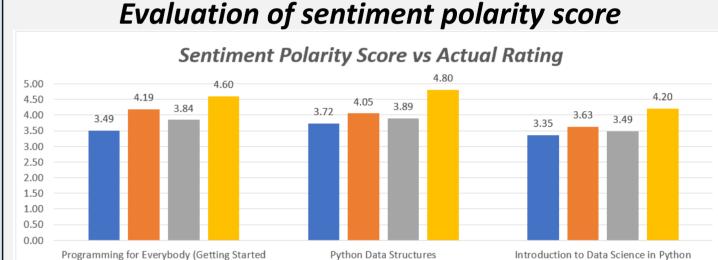
	Programming for Everybody (Getting Started with Python)	Python Data Structures	Introduction to Data Science in Python
bigram	charles severance  dr Chuck easy understand basic python learn basic earn python course really want learn course really want learn highly recommend course agreat course take coursestart python good course programming language	good course learn python great course learn lot course learn lot dr chuck course learn data structure nice course easy understand python data course really course beginner	discussion forum  data science learn lot use python use python scourse assignment course really great course excellent course data analysis good course course learn course good python data
trigram	charles russell severance learn basic python really enjoy course good course beginner start learn python highly recommend course course dr chuck dr charles severance dr chuck dr charles severance want learn python great course beginner start learn python highly recommend course course dr chuck dr charles severance want learn python great course beginner get start python great course beginner	data structure data structure  data structure python  or chuck great  great course Isann  or charles beverance  thank dr chuck thank dr chuck learn lot course  course learn python  learn data structure	really good course field data science data science course spend lot time data science use python learn data science great course learn course data science good course learn lot learn lot course one best course one best course data science python data analysis start data science python data science python data science python data science use python data basic data science of the data

# **Sentiment Analysis**

	Average of Polarity		
name	Textblob	Vader	Overall
Programming for Everybody (Getting Started with Python)	0.40	0.68	0.54
Python Data Structures	0.49	0.62	0.55
Introduction to Data Science in Python	0.34	0.45	0.40

	Programming for Everybody (Getting Started with Python)	Python Data Structures	Introduction to Data Science in Python
bigram	course would want learn want learn dr chuck learn basic course beginner basic python basic python complete course learn python waste time look forward good course	course bit  learn lot dr chuck previous one well explain  data structure  unnecessary talk previous course  bit difficult  complete assignment  data set solve problem new thing bit tough  little bit python programming first course course material	waste time assignment difficult take course course content course assignment good course complete assignment data science auto grader learn lot stack overflow assignment reallylot time bad course wersion panda course material discussion forum complete course video lecture
trigram	take long time get start python     waste time learn basic python program  apply financial aid     write basic code     run wrong program     course would helpful learn basic python want learn program     wnt learn python long time figure  want learn python	talk problem set data structure course much irrelevant information guy talk problem data structure python practice need help teaching guy talk tough previous one set practice need set practice need set practice need set practice need great course give need help pause little bit tough pause video try	data science course ever take coursera python data science take much long assume know everything would recommend course course much good spend lot time read discussion forum machine learn course bad course ever assignment grade system assignment really difficult waste much time waste time money real world problem

### Discussion and Finding Evaluation of accuracy and usability of n-gram dr chuck data structure good course good course easy understand excellent course excellent course discussion forum course really easy understand course good course learn python data data science pythor python data science roduction data scien get start python learn data science start learn python good course beginner dr chuck great course data science thanks dr chuck dr charles severance learn lot course nighly recommend course learn python data course learn lot start data science really enjoy course



harles russell severance

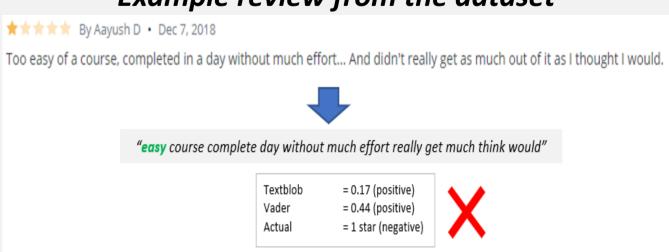
### Confusion matrix and classification report

		Predicted negative			Predicted positive			
Observed nego	bserved negative		326		385			
Observed posi	Observed positive		772		28324			
	precision recall f		f1	-score	support			
negative	3	30%	46%		36%	711		
positive	9	99%	97%		98%	29096		
accuracy			96%		29807			
		- 40/	700/		c=0/	2222		

### Example review from the dataset

96%

97%



### Conclusion

The n-gram analysis and word cloud are sufficient enough to provide an accurate and informative glance into the course. However, it falls short on sentiment analysis especially in detecting the negative reviews.