Prerequisites

In order to be successful in this course, you will need to know how to program in Python. The expectation is that you have completed the first three courses in this Applied Data Science with Python series, specifically Course 1 on Introduction to Data Science in Python and Course 3 on Applied Machine Learning in Python, so that you are familiar with the numpy and pandas Python libraries for data manipulation, and scikit-learn toolkit for machine learning algorithms.

Week by week

In this course, you will build your skill sets in text mining through lectures on core topics, code walk-throughs on tasks related to weekly assignments, reading material, quizzes, and programming assignments.

In **Module One**, you will be introduced to basic text mining tasks, and will be able to interpret text in terms of its building blocks – i.e. words and sentences, and reading in text files, processing text, and addressing common issues with unstructured text. You will also learn how to write regular expressions to find and extract words and concepts that follow specific textual patterns. You will be introduced to UTF-8 encoding and how multi-byte characters are handled in Python. This week's assignment will focus on identifying dates using regular expressions and normalize them.

In **Module Two**, you will delve into NLTK, a very popular toolkit for processing text in Python. Through NLTK, you will be introduced to common natural language processing tasks and how to extract semantic meaning from text. For this week's assignment, you'll get a hands-on experience with NLTK to process and derive meaningful features and statistics from text.

In **Module Three**, you will engage with two of the most standard text classification approaches, viz. naïve Bayes and support vector machine classification. Building on some of the topics you might have encountered in Course 3 of this specialization, you will learn about deriving features from text and using NLTK and scikit-learn toolkits for supervised text classification. You will also be introduced to another natural language challenge of analyzing sentiment from text reviews. For this week's assignment, you will train a classifier to detect spam messages from non-spam ("ham") messages. Through this assignment, you will also get a hands-on experience with cross-validation and training and testing phases of supervised classification tasks.

In **Module Four**, you will be introduced to more advanced text mining approaches of topic modeling and semantic text similarity. You will also explore advanced information extraction topics, such as named entity recognition, building on concepts you have seen through Module One and Module Three of this course. The final assignment lets you explore semantic similarity of text snippets and building topic models using the gensim package. You will also experience the practical challenge of making sense of topic models in real life.

Enrollment Options

Coursera has made the decision to make Specializations available by monthly subscription. This means you can choose to pay a monthly fee to access all of the courses in a specific Specialization.

Coursera's switch to monthly subscriptions comes with another change -- for those learners who choose the "Audit Only" enrollment, you will no longer be able to submit assignments for grades nor see answers for those assignments. You will still have access to all the course materials but you will not be graded on your work, nor see answers to graded assignments.

For further information on the different enrollment options for Coursera courses, please visit the Enrollment Options Help page. If you have feedback about the enrollment options shared on the Enrollment Options page, you can share your thoughts with Coursera in this survey.

Grading and Assignments

The lectures will provide you with some guidance for completing assignments, but you will need to take initiative and look beyond assignment instructions in order to be successful. You'll need to know how to ask questions in the discussion forums of your peers, and seek out new information through web searches and Stack Overflow. Be sure to also check out the Additional Resources.

If you are not sure what kind of output is required, or think there is a need for more clarity, please head to the course discussion forums. Note that some assignments and in video quizzes may not be mobile friendly.

Some assignments allow you to download and view your fellow learner's code and/or data. If you want to look at the learner's code, we recommend that you open it through the Jupyter notebook system on the Coursera platform as that will be more secure. Please ensure that all data you share is publicly available, since you will be sharing these data with other learners.

Course Item	**Percentage of Final Grade**	**Passing Threshold**
Week 1 Quiz	5%	80%
Week 1 Jupyter Notebook Assignment	20%	
Week 2 Quiz	5%	80%

Week 2 Jupyter Notebook Assignment	20%	
Week 3 Quiz	5%	80%
Week 3 Jupyter Notebook Assignment	20%	
Week 4 Quiz	5%	80%
Week 4 Jupyter Notebook Assignment	20%	

Code of Conduct

Visit Coursera's Code of Conduct and to abide by guidelines there. It is important when giving feedback to your peers to be polite and to be sensitive to the diversity of cultures and backgrounds of learners in your course.

Working Offline

While the Coursera platform has an integrated Jupyter Notebook system, you can work offline on your own computer by installing Python 3.5+ and the Jupyter software packages. For more details, consult the Jupyter Notebook FAQ.

Note that this course uses the following packages: (a) sklearn-0.18.1, (b) nltk-3.2.4, (c) gensim-2.1.0, and (d) pandas-0.20.1.

Accessibility

We strive to develop fully accessible courses. Occasionally, some of our content does not fully meet our accessibility goals. Please use this form to inform us of any accessibility issues you are experiencing in this course.

Help!

If you're having problems, here are a couple of great places to go for help:

If the problem is with the Coursera platform such as verification on assignments, in video quiz problems, or the Jupyter Notebooks, please check out the Coursera Learner Support Forums.

If the problem deals with understanding the assignment or how to use the Jupyter Notebooks, please read our Jupyter Notebook FAQ page in the course resources

If you have questions with the content of the course, or questions about programming in python or with the toolkits described, you can contact your peers and the course instructors in the discussion forums, or go to Stack Overflow.

Having trouble accessing your previously submitted assignments? If your session has ended, you can access these again by selecting the "Switch Session" option. Details for how to select this can be found in this learner help center article. If you still have issues accessing your materials after switching sessions, please reach out to Coursera learner support via our online chat forums in the Learner Help Center.

In-Video Questions (IVQs)

In this course, in-video questions or IVQs may appear during lectures to help you learn as well as assess your understanding of the content. IVQs are optional and do not count towards your overall course grade.

Types of in-video questions

Many of the lectures contain in-video questions (IVQs). These questions are presented in a variety of formats. Some will ask you to write or think about a concept from the video. Others will ask for a short answer. Still others may ask you to choose from a multiple-choice list of answers. If an IVQ is a survey or a poll, you will see a summary of responses from other learners after you respond. You can look at the question again later to see new summary data as more of your peers answer.

Some IVQs also contain runnable code blocks. These IVQs allow you to practice the coding concepts during the lecture. In this course, these types of IVQs will usually be directly followed with the solution code.