

Course Description

In this course, you will get to know one important subfield in data mining: pattern discovery. You will learn why pattern discovery is important, what the major tricks are for efficient pattern mining, and how to apply pattern discovery in some interesting applications. This course provides you the opportunity to learn concepts, principles, and skills to practice and engage in scalable pattern discovery methods on massive data; discuss pattern evaluation measures; study methods for mining diverse kinds of frequent patterns, sequential patterns, and sub-graph patterns; and explore their applications.

Course Goals and Objectives

By the end of this course, you will be able to:

- Recall important pattern discovery concepts, methods, and applications, in particular the basic concepts of pattern discovery, such as frequent pattern, closed pattern, max-pattern, and association rules.
- Identify efficient pattern mining methods, such as Apriori, ECLAT, and FPgrowth.
- Compare pattern evaluation issues, especially several popularly used measures, such as lift, chi-square, cosine, Jaccard, and Kulczynski, and their comparative strengths.
- Compare mining diverse patterns, including methods for mining multi-level, multi-dimensional patterns, qualitative patterns, negative correlations, and compressed and redundancy-aware top-k patterns.
- Learn well-known sequential pattern mining methods, including methods for mining sequential patterns, such as GSP, SPADE, PrefixSpan, and CloSpan.
- Enjoy various pattern mining applications, such as mining spatiotemporal and trajectory patterns and mining quality phrases.
- Explore further topics on pattern analysis, such as pattern mining in data streams, software bug mining, pattern discovery for image analysis, and privacy-preserving data mining.

If you have additional time and would like to go deeper on certain advanced pattern discovery topics, you may like to delve into some optional, supplementary materials associated with this course, such as:

- Constraint-based pattern mining, including methods for pushing different kinds of constraints, such as data and pattern-based constraints, anti-monotone, monotone, succinct, convertible, and multiple constraints.
- Mining long (colossal) patterns.
- Graph pattern mining, including methods for subgraph pattern mining, such as gSpan, CloseGraph, graph indexing methods, mining top-k large structural patterns in a single large network, and graph mining applications, such as graph indexing and similarity search in graph databases.
- Pattern-based classifications, including CBA, CMAR, PatClass, DDPMine, and DPClass.

Textbook and Readings

Although the lectures are designed to be self-contained, we recommend (but do not require) that you refer to the following book:

Han, Kamber, and Pei. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 3rd ed., 2011.

You can download a [PDF version of Chapters 1, 6, and 7](#) from *Data Mining: Concepts and Techniques*, 3rd ed. for free.

If you would like to purchase the entire textbook, the publisher has an exclusive offer just for Coursera students. You can save 30% on either the print or eBook version of *Data mining: Concepts and techniques* (3rd ed.) and receive free shipping on all orders. Here is how it works:

1. [Add the book to your cart.](#)
2. Enter code **COMP315** and click **Apply**.
3. The discount will be applied to the list price and cannot be combined with other promotions.

Additional, optional materials may have been released or will be released shortly, and you can study them based on your own interest and proceed according to your own pace.

Course Outline

The course consists of **4 weekly modules, with each week consisting of 2 lessons**.

Please note: There are no required readings for this course. All readings listed below are optional.

| Module | Key Concepts | Suggested Readings from <i>Data Mining: Concepts and Techniques</i> |
|--------|--|--|
| Week 1 | <ul style="list-style-type: none"> • Pattern Discovery: Basic Concepts • Efficient Pattern Mining Methods | <ul style="list-style-type: none"> • Chapter 6: Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods (Sections 6.1-6.2) |
| Week 2 | <ul style="list-style-type: none"> • Pattern Evaluation • Mining Diverse Patterns | <ul style="list-style-type: none"> • Chapter 6: Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods (Section 6.3) • Chapter 7: Advanced Pattern Mining (Sections 7.1-7.5) |
| Week 3 | <ul style="list-style-type: none"> • Sequential Pattern Mining • Pattern Mining Application: Spatial and Trajectory Pattern Mining | <ul style="list-style-type: none"> • Chapter 13: Data Mining Trends and Research Frontiers (Sections 13.1-13.2) • Some recent research papers listed in the lecture videos/slides |

| | | |
|---------------|--|---|
| Week 4 | <ul style="list-style-type: none"> • Pattern Mining Applications: Mining Quality Phrase from Text Data • Advanced Topics on Pattern Discovery | <ul style="list-style-type: none"> • Some recent research papers listed in the lecture videos/slides |
|---------------|--|---|

Elements of This Course

The course is comprised of the following elements:

- **Lecture videos:** Each week your instructor will teach you the concepts you need to know through a collection of short video lectures. You may either stream these videos for playback within the browser by clicking on their titles, or you can download each video for later offline playback by clicking the download icon.
- **In-lecture questions:** Some lectures have questions associated with them to help verify your understanding of the topics. These questions will automatically appear while watching the video if you stream the video through your browser. The questions are available as a separate downloadable text file for those who prefer to download the videos. These questions do not contribute toward your final score in the class.
- **Quizzes:** Each lesson will include one graded quiz. Your cumulative score will be used when calculating your final score in the class. There is no time limit on how long you take to complete each quiz. The deadline for all quizzes is the last day of the course.
- **Programming assignments:** There are two programming assignments for this class. The first programming assignment is about understand and implementation of the most well-known frequent pattern mining algorithm: Apriori, based on some real-life dataset; and the second programming assignment is about understanding and implementation of some simple method for contiguous sequence pattern mining and phrase mining, based on some real-life dataset. The first programming assignment is required for the Course Certificate. The second programming assignment is optional.

Information about Lectures

The lectures in this course contain the most important information you need to know. You can access these lectures in each week's lesson section. The following resources accompany each video:

- The play button will open the video up in your browser window and stream the lecture to you. The duration of the video (in hours-minutes-seconds format) is also listed. Some lectures may include in-video questions described above. Within the player that appears, you can click the CC button to activate closed captions. English captions are available for all videos.
- All video lectures have a discussion forum dedicated to them. This is a great place to discuss any questions you have about the content of the video or to share your ideas and responses to the video.

Discussion Forums

The discussion forums are a key element of this course. Be sure to read more [about the discussion forums](#) and how you can make the most of them in this class.

How to Pass This Course

I am continually looking to improve this course and may encounter some issues requiring us to make changes sooner rather than later. As such, this syllabus is subject to change. I appreciate your input and ask that you have patience as we make adjustments to this course.

To qualify for a Course Certificate, simply start verifying your coursework at the beginning of the course, get a **75% or higher on all graded lesson quizzes**, and **70% or higher on the required programming assignment**, and pay the fee. Coursera [Financial Aid](#) is available to offset the registration cost for learners with demonstrated economic needs. If you have questions about Course Certificates, [please see the help topics here](#).

Also note that this course is in the [Data Mining Specialization](#) offered by the University of Illinois at Urbana-Champaign. By earning a Course Certificate in this course, you are on your way toward earning a [Specialization Certificate in Data Mining](#). You may also choose to pre-pay for the entire Specialization, at a discount. See more information about [Specialization payments](#) here.

If you choose not to pay the fee, you can still audit the course. You will still be able to view all videos, submit practice quizzes, and view required assessments. Auditing does not include the option to submit required assessments. As such, you will not be able to earn a grade or a Course Certificate.

Getting and Giving Help

You can get/give help via the following means:

- Use the [Learner Help Center](#) to find information regarding specific technical problems. For example, technical problems would include error messages, difficulty submitting assignments, or problems with video playback. If you cannot find an answer in the documentation, you can also report your problem to the Coursera staff by clicking on the **Contact Us!** link available on each topic's page within the Learner Help Center.
- Use the [Content Issues](#) forum to report errors in lecture video content, assignment questions and answers, assignment grading, text and links on course pages, or the content of other course materials. University of Illinois staff and Community Mentors will monitor this forum and respond to issues.

Note: Due to the large number of learners enrolled in this course, I am not able to answer emails sent directly to my account. Rather, all questions should be reported as described above.