Course Code
Big Data Analytics
Module 4, Spring 2019

Course Information

Instructor:
Office: PHBS Building, Room 708
Phone: 86-755-2603- Mobile: 18319058199
Email: lzwqhk@139.com
Office Hour: 920-1020 Monday/Thursday

Teaching Assistant:
Phone:
Email:

Classes:
Lectures: Monday/Thursday, 10:30am-12:20am
Venue: PHBS Building,

Course Website:
If any.

1. Course Description

1.1 Context

Course overview:
This course intends for FinTech graduate students to gain fundamental understanding of big data analytic (BDA) techniques and algorithms for finance. Towards this objective, this course is designed in such a way that lectures on BDA will go hand in hand with financial system development and applications. With this setup, in the beginning of the course, introductory materials about how big data analytics impact financial service industry will be provided. Students enrolled are expected to have grasped at least basic programming skills.
This course provides a general introduction to classic big data analytic techniques and algorithms which will be made and present relevant for finance. BDA topics include MapReduce, Principle Component Analysis (PCA), Singular Value Decomposition (SVD), locality sensitive hashing/finding similar items, data stream processing, link analysis and mining social networks, clustering, frequent itemsets, advertising on the web, recommendation systems, regression and large scale machine learning. Students in this course will be required to complete a project on developing an intelligent financial asset analysis platform applying the BDA techniques and algorithms taught.

Prerequisites:

Programming skills and basic statistics and probabilities required.

Data structures training preferred.

1.2 Textbooks and Reading Materials

Textbook for this course: Mining of Massive Data Sets, 2nd Edition, Cambridge University Press. Other readings recommended are Python for Finance, Analyzing Big Financial Data, Yves Hilpisch, O'Reilly.

2. Learning Outcomes

2.1 Intended Learning Outcomes

Learning Goals	**Objectives**	**Assessment (YES with details or NO)**
1. Our graduates will be effective communicators.	1.1. Our students will produce quality business and research-oriented documents.	Students are assessed in conducting BDA projects.
	1.2. Students are able to professionally present their ideas and also logically explain and defend their argument.	Students are assessed in presenting their project results.
2. Our graduates will be skilled in team work and leadership.	2.1. Students will be able to lead and participate in group for projects, discussion, and presentation.	Students are assessed in their role in the BDA projects.
	2.2. Students will be able to apply leadership theories and related skills.	Students are assessed about their performance in their project.
3. Our graduates will be trained in ethics.	3.1. In a case setting, students will use appropriate techniques to analyze business problems and identify the ethical aspects, provide a solution and defend it.	Students are assessed in their projects which will include ethics.
	3.2. Our students will practice ethics in the duration of the program.	Students are assessed in their projects which will include ethics.
4. Our graduates will have a global perspective.	4.1. Students will have an international exposure.	Students are assessed about their performance in their project.
5. Our graduates will be skilled in problem-solving and critical thinking.	5.1. Our students will have a good understanding of fundamental theories in their fields.	Students are assessed about their performance in their project.
	5.2. Our students will be prepared to face problems in various business settings and find solutions.	Students are assessed about their performance in their project.
	5.3. Our students will demonstrate competency in critical thinking.	Students are assessed about their performance in their project.

2.2 Course specific objectives

Students who complete this course should be able to understand the strengths and weaknesses of certain big data analytic techniques and algorithms. Students should be able to apply this understanding to select appropriate big data analytics techniques and algorithms in analysing massive financial data.

2.3 Assessment/Grading Details

Assessment will be conducted based on their project performance, lecture discussion participation, and lecture quizzes. Grading will be based on the discussion participation, report submitted, project execution, and their project presentation. Grading details are as follows:

- Lecture discussion participation: 5%
- Ethical issues: 5%
- Lecture quizzes: 20%
- Project on developing an intelligent financial asset analysis platform (iFAA): 70%
    - Big data analytic algorithms implementation: 30%
    - Financial asset trading strategies implementation: 20%
    - Final project report, demonstration, and presentation: 20%

2.4 Academic Honesty and Plagiarism

It is important for a student's effort and credit to be recognized through class assessment. Credits earned for a student work due to efforts done by others are clearly unfair. Deliberate dishonesty is considered academic misconducts, which include plagiarism; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.

All assessments are subject to academic misconduct check. Misconduct check may include reproducing the assessment, providing a copy to another member of faculty, and/or communicate a copy of this assignment to the PHBS Discipline Committee. A suspected plagiarized document/assignment submitted to a plagiarism checking service may be kept in its database for future reference purpose.

Where violation is suspected, penalties will be implemented. The penalties for academic misconduct may include: deduction of honour points, a mark of zero on the assessment, a fail grade for the whole course, and reference of the matter to the Peking University Registrar.

For more information of plagiarism, please refer to *PHBS Student Handbook*.

3. Topics, Teaching and Assessment Schedule

Tentative course topics and content are listed in the following table, which will be adapted to the situation of the class.

Week	Course content	Course topics
1st week	Introduction to Big Data Analytics for Finance	• Typical financial systems and financial service innovation • Statistical modeling and machine learning • Big data analytics and its impact • MapReduce • Algorithms using MapReduce • Extensions to MapReduce • The communication cost model • The complexity theory for MapReduce
2nd week	Finding similar items	• Application of set similarity

		• Shingling of documents • Similarity-preserving summaries of sets • Locality-sensitive hashing for documents • Distance measure
3rd week	Data stream processing	• The stream data model • Sampling data in a stream • Filtering streams • Counting distinct elements • Estimating moments • Counting ones in a window
4th week	Link analysis and mining social networks	• PageRank • Efficient computation of PageRank • Topic sensitive PageRank • Link Spam • Hubs and authorities • Social networks as graphs • Clustering of social networks • Community discovery
5th week	Frequent itemsets	• The market-basket model • Market basket and A-Priori algorithm • Handling larger datasets in main memory • Limited-pass algorithms • Counting frequent items in a stream
6th week	Clustering	• Introduction to clustering • Hierarchical clustering • K-Means algorithms • The CURE algorithm • Clustering in non-Euclidean spaces • Clustering for streams and parallelism
7th week	Advertising on the web	• Issues in online advertising • Online algorithms • The matching problem • The Adwords problem • Adwords implementation
8th week	Recommendation systems	• A model for recommendation systems • Content based recommendations • Collaborative filtering • Dimensionality reduction • Eigenvalues and eigenvectors • PCA and SVD • CUR decomposition • Netflix challenge
9th week	Regression, large scale machine learning, and other BDA topics	• Machine learning models • Perceptrons • Support-vector machine • Learning from Nearest neighbors • Decision trees • Deep learning • Regression and optimization • Game theory • Interpretive structural modeling (ISM) and structural eruptional modeling (SEM)

4. Miscellaneous

- Student team. Students will be divided into small teams with size of about 3.
- Lectures. Classes meet twice each week.
- Quiz: Quiz will be conducted during lectures to test students' preparation for as well as grasping of the course content.
- Presentation sessions. Each team will be required to present their progress on the project using time slots allocated within lectures.
- Communication: It is encouraged to communicate via email and face to face meeting during office hours.
- Readings and reference book. We will use textbook Mining Massive Data Sets, 2nd Edition, Cambridge University Press, as reference book. Other readings recommended are Python for Finance, Analyzing Big Financial Data, Yves Hilpisch, O'Reilly.
- Project and grading: Each student team need to complete a project.
- Data sets: financial asset transaction data sets will be provided.