Course Introduction and Syllabus

Zhejiang University

Course code：2021051 **Course title：Unstructured Data Analytics**

Credit：2 **Teaching hour per week：3.5 hours/week**

Target students： Master students

Instructor: Dr. Kong Xiangwei **Email: kongxiangwei@zju.edu.cn**

Prerequisites：Python\ Calculus \ Linear algebra \ Statistics

I. Course Introduction

This course introduces students to the process of performing high-valued analytics using unstructured data to support business decisions. Unstructured data (text, tweets, posts, video, audio, etc.) consist of over 80% of the data being produced every day. Unstructured data analysis includes methods of text mining, image classification, video analysis, music recommendation, etc., and its application permeates in commercial, manufacturing, financial, medical and daily life. Most companies have minimal competencies in using unstructured data analytics for new product development, customer retention, workforce optimization, and a myriad of other areas.so this course will allow students to begin to master how to make decisions based on unstructured data, through basic unstructured data analysis and typical applications, Students can master the most advanced data analysis theory methods and models, combined with the industry big data to carry out analysis, promote data analysis and management engineering development. This course will get you started in how to lead efforts at your company to monetize their data and achieve your career objectives.

II. Teaching Objectives

i. Learning Objectives

By the conclusion of the course of study, students are expected to be able to:

1) Know how to use unstructured data analysis methods and models in business to solve problems and make decisions;
2) Understand the key methods and model principles in text analysis, image analysis, video analysis and audio analysis;
3) Understand the typical applications of unstructured data analysis in commerce, manufacturing, finance, medical treatment, and people's livelihood;
4) Know the most advanced data analysis theoretical methods and models, carry out analysis under industry big data, and promote the development of data analysis and management engineering disciplines.

ii. Measurable Learning Outcomes

Expected learning outcomes for students include:

1) Utilize basic unstructured data analysis models and analyze applications in typical industries or competitions;
2) Utilize programming to build and implement 1-2 unstructured data analysis models under the big data application scenarios.

III. Course Requirements

i. Teaching Methods and Requirements

1) Every week students must attend one 3.5-hour lecture.
2) **Lectures** will be given to present the key conceptual material through discussion and interaction between lecturer and students. Lectures are supported by readings, class discussions and illustrations of real-world case examples.
3) **Guest speaker** may be invited to give lectures/seminars on specific issues related to the social media marketing or digital platform industries to enhance students' understanding of the theories presented in lecture.
4) **Course Calendar** The course calendar (as presented below) details scheduling information. Note that this calendar may change as the course proceeds. Any changes will be announced.

Students are expected to prepare for and attend all classes to gain full benefit from the course

ii. Course Evaluation and Grading

All material presented is examinable (except where stated otherwise) by assignments and the final examination. All important assessment information such as due dates and times, content, guidelines and so on will be discussed at lectures. Students are responsible for ensuring that they are aware of this information, keeping track of their own progress, and catching up on any missed classes.
Students are required to conduct a case study about a real-world case or a competition case on platforms such as Kaggle, which use unstructured data.
Students are required to complete an unstructured data analysis project. Students will use data like text, image and use state-of-the-art models talked during lectures. Programming, a final report and a final representation are required.

Assessment	**% of final grade**
Class participation	5%
Case study	15%
Final assignment and presentation	75%

Extensions can only be given by your lecturer and only in special circumstances such as:

- Illness: a medical certificate is required
- family emergency
- representative activities (sport, cultural, academic etc.)

Verification of circumstances is required to validate extensions.

Computer problems are not deemed to warrant extensions, except in the situation of Zhejiang University's computer system failure. Students should ensure they allow sufficient time to overcome these problems before the assignment is due. **Students are expected to keep hard copies, draft and backup files of work done.** These can then be used in cases of computer failure, plagiarism, and in other circumstances as required. **Full academic workloads and work commitments are not deemed to warrant extensions**, as students are aware of requirements at the beginning of semester as featured in syllabus.

IV. Teaching Schedule

Lecture	Topic
1	Course Overview; Introduction to Unstructured Data Analytics
2	Unstructured Data Source and Representation; Basic Concept and Feature Extraction
3	Text Analysis: Bag of words, TF, DF, IDF etc.; Advanced Topics: Basic Text Analysis and Mining
4	Image Analysis (SURT, SIFT, HOG etc.); Advanced Topics: Finding Poverty in Satellite Image
5	Image Analysis with CNNs; Advanced Topics: Image Classification and Retrieval for Products
6	Video Analysis; Advanced Topics: Rid for Self-service Supermarket
7	Audio Analysis (MFCC, Style Analysis etc.); Advanced Topics: Musical Instrument Identification
8	Final Presentation

V. References and Recommended Readings

1. C. Zhai and S. Massung, Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining, ACM Book Series, Morgan & Claypool Publishers, 2016.
2. Richard Szeliski,Computer Vision: Algorithms and Applications, Springer, 2011th Edition, 2010
3. Ian, Goodfellow,Yoshua,Bengio,Aaron. Deep Learning, MIT Press,2016.