

Natural Language Processing in Business Using Deep Learning

ISOM4000B: Fall 2021

Course Name	Natural Language Processing in Business Using Deep Learning
Course Code	ISOM 4000B (3 Credits)
Exclusion	COMP 4221, COMP 4901K, MATH 4824B, ELEC 4230
Prerequisite	ISOM 3360 or ISOM 3400
Instructor	Dr. Boris K C NG
Contact	Office: LSK 4016A Office Hours: By Appointment Email: borisng@ust.hk Begin subject: [ISOM4000B]
Course Schedule and Classroom	Lecture: Mon: 1:30pm - 3:20pm Lab: Thu: 5:00pm - 7:50pm
Course Webpage	Accessible from Canvas

Course Overview

The ability to process unstructured data has become an essential capability of a company. User inputs are no longer restricted to itemized menus or structured forms based on a pre-defined logic flow. Increasingly, firms need to interface directly with data sources with no pre-defined structure and to convert unstructured data into signals and useful inputs to the applications. The concept of text as data underlies many business applications that interface with human subjects. Social media applications are obvious examples but financial technologies (fintech) applications also exemplify this trend of relying on unstructured text processing in delivering novel financial services to customers.

Advance in business intelligence and fintech has given rise to techniques to decipher business documents and user-generated content. Lying at the core of these techniques is the natural

language processing (NLP) model used to filter textual inputs to extract signals that affect customer behaviors and business outcomes. Major source of unstructured inputs to these systems can be news, commentaries, and reports extracted from social media, traditional news media and information vendors.

This course covers the concepts and techniques for applying natural language processing (NLP) and deep learning in processing text data in the business domain. Python programming language will be the primary implementation tool for data analysis and model development. An extended lab session intends to train students with good knowledge and skills for solving practical problems, such as document classification, word embedding, topic modeling, and financial news analysis.

After taking this course, you should:

1. Know the basics of NLP techniques and deep learning models well enough to interact with business analysts, data scientists, engineers, and machine learning developers.
2. Be able to apply NLP techniques and deep learning approaches to analyze text for business intelligence.
3. Gain hands-on experience in using popular NLP and deep learning tools and get ready for the job positions that require familiarities with the NLP and deep learning tools.

The detailed course schedule is shown below:

Week	Date	Topics	Assignments
1	Sep 6	C1 - Course Overview	
	Sep 9	LAB1 - Review on Python, Data Preparation, and Machine Learning Implementations	
<i>Introduction to Natural Language Processing and Deep Learning</i>			
2	Sep 13	C2 - Introduction to Natural Language Processing I	HW1 release
	Sep 16	LAB2 - Basic NLP Implementations	
3	Sep 20	C3 - Introduction to Natural Language Processing II	
	Sep 23	LAB3 - Lexicon, Topic Modeling, and Knowledge Graph	
4	Sep 27	C4 - Introduction to Deep Learning I	
	Sep 30	LAB4 - Neural Networks and Multilayer Perceptron	
5	Oct 4	C5 - Introduction to Deep Learning II	HW1 due
	Oct 7	LAB5 - Deep Autoencoders	
<i>Combining Natural Language Processing with Deep Learning</i>			
6	Oct 11	C6 - Limitations in Conventional NLP Approaches	HW2 release
7	Oct 18	C7 - Word Embedding Model	
	Oct 21	LAB6 - Word2vec	
8	Oct 25	C8 - Word Embeddings in Supervised Learning	
	Oct 28	LAB7 - CNN and RNN	
9	Nov 1	C9 - Word Embeddings in Unsupervised Learning	HW2 due
	Nov 4	LAB8 - Word Similarity and Its Applications	
<i>Applications in Business</i>			
10	Nov 8	C10 - Computerized Textual Analysis of Fake News	Project release
	Nov 11	LAB9 - Deep Learning and Econometric Analyses	
11	Nov 15	C11 - Conversational AI Chatbots	
	Nov 18	LAB10 - Developing a Chatbot	
12	Nov 22	C12 - NLP and Ethics in Recruitment	
	Nov 25	LAB11 - Attention, Transformer, and BERT	
13	Nov 29	Final Examination	Project due

Lecture Notes and Readings

- All course materials (Lecture slides, assignments, and lab handouts) are available on the class website.
- Supplemental books (optional):
 1. *Natural Language Processing with Python*, by Steven Bird, Ewan Klein & Edward Loper, O'Reilly Media, 2009 ISBN: 978-0-596-51649-9
 2. *Deep Learning for Natural Language Processing*, by Palash Goyal, Sumit Pandey & Karan Jain, Apress Media, 2018 ISBN: 978-1-4842-3684-0

Grading

Your grades will be determined based on homework assignments, an individual project and a final examinations.

Homework Assignments ($\times 2$)	30% (15% $\times 2$)
Project	40%
Final Exam	30%
Total	100%

Important Notes on the Lab Session

This course emphasize the lab sessions as an essential part of the learning process in the form of active practice. You are NOT going to learn without practicing the data analysis yourselves. During the lab session, I will expect you to be entirely devoted to the class by following the instructions. And you should actively link the empirical results you obtained during the lab to the concepts you learned in the lectures. Besides, I will deliver the necessary knowledge and skills for you to complete the homework assignments and individual project. Although I will not take attendance for the lab participation, you are expected to attend all labs in order to maximize the learning outcomes.

Important Notes on the Class Participation

I highly appreciate your in-class participation. I will expect you to actively ask questions and participate in group discussions.

Homework Assignment, Project and Final Exam

- Homework Assignment (15% × 2)

There will be a total of **two** individual homework assignments, each comprising questions to be answered and hands-on tasks. In the homework, you will gain experience on how to apply NLP techniques and deep learning in *Social Media Monitoring*. We will leverage a Twitter dataset containing tweets about COVID-19 to study how individuals response to the pandemic and how they engage with COVID-19 tweets with different sentiment and emotions. The purpose of the homework is to provide step-by-step instructions on social media text mining and sentiment analysis using NLP and deep learning.

Completed assignments must be handed in via Canvas **prior to the start of the class** on the due date. Assignments will be graded and returned promptly.

Turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date. Assignments up to 24 hours late will have their grade reduced by 25%; assignments up to one week late will have their grade reduced by 50%. After one week, late assignments will receive no credit.

- Project (40%)

There will be **one** individual project. In this project, you will apply the NLP techniques and deep learning models you learned in the class to solve a real-world problem. More specifically, you need to decipher financial news articles for *News-Induced Market Signaling*. The deliverable is a written report summarizing what you have done and what you have achieved. More details will be provided later.

- Final Exam (30%)

The final exam is on Monday, **Nov 29**, in classroom. More details on the final exam will be provided later.

Academic Integrity

Students at HKUST are expected to observe the Academic Honor Code at all times (see <http://acadreg.ust.hk/generalreg.html> for more information). Zero tolerance is shown to those who are caught cheating on any quiz or exam. In addition to receiving a zero mark on the quiz or exam involved, the final course grade will appear on your record with an X, to show that the grade resulted from cheating. This X grade will stay with your record until graduation. If you receive another X grade, you will be dismissed from HKUST.