CS 6501 Natural Language Processing

Introduction

Yangfeng Ji

Information and Language Processing Lab
Department of Computer Science
University of Virginia

ENGINEERING

About Lecture Format

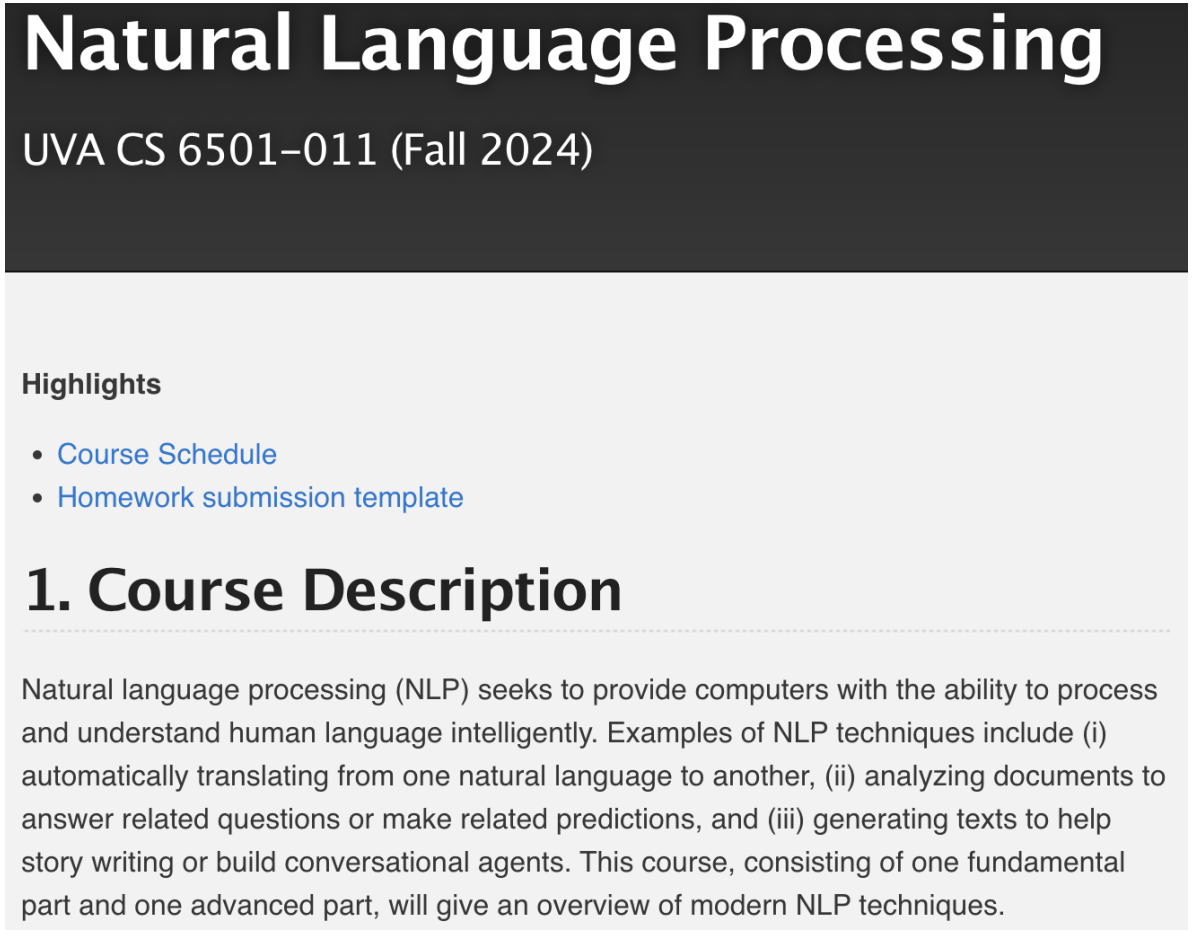
- All lectures will be given in Olsson 005, and *recorded* and uploaded to Canvas
  - No online participation is allowed
- We use Piazza for online discussion
  - In-class and in-person QA are highly encouraged

Classroom Policy

- Computers
  - You can use your computer during lecture to take notes
- Attendance
  - We do not take attendance in this class
  - In-person participation is more than necessary
- QA
  - Feel free to ask questions during lectures

Course Information

Course Webpage



Natural Language Processing

UVA CS 6501-011 (Fall 2024)

Highlights

- Course Schedule
- Homework submission template

1. Course Description

Natural language processing (NLP) seeks to provide computers with the ability to process and understand human language intelligently. Examples of NLP techniques include (i) automatically translating from one natural language to another, (ii) analyzing documents to answer related questions or make related predictions, and (iii) generating texts to help story writing or build conversational agents. This course, consisting of one fundamental part and one advanced part, will give an overview of modern NLP techniques.

https://yangfengji.net/uva-nlp-grad/

Instructors

- Instructor
  - Yangfeng Ji
  - Office hour: Wednesday 11 AM - 12 PM
  - Location: Rice 510

Instructors

- Instructor
  - Yangfeng Ji
  - Office hour: Wednesday 11 AM - 12 PM
  - Location: Rice 510
- TAs:
  - Caroline Gihlstorf (Office hour: Monday 11 AM - 12 PM, Rice 414)
  - Elizabeth Palmieri (Office hour: Tuesday 2 - 3 PM, Rice 414)
  - Nibir Chandra Mandal (Office hour: Thursday 2 - 3 PM, Rice 414)

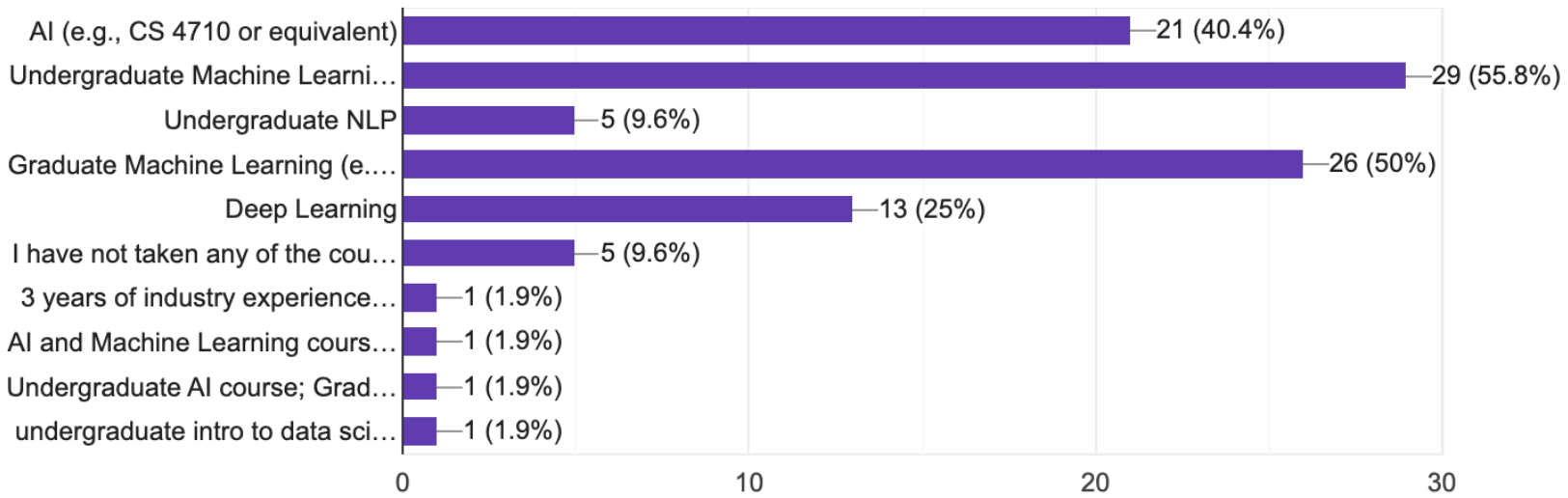
Survey Results: Background

Background: Learning NLP requires a good understanding of machine learning/deep learning.

Copy

Please select all the courses that you have taken before (not in Fall 2024).

52 responses



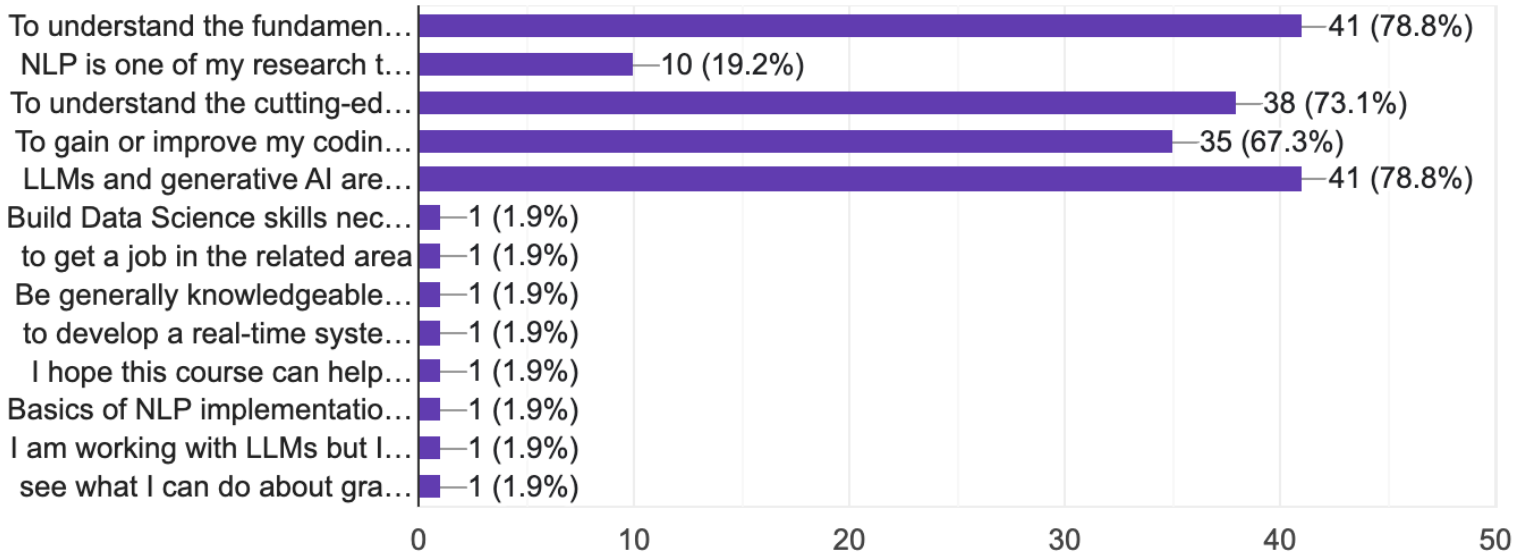
Survey Results: Learning objectives

Learning Objective: What are the objectives of taking this class?

Copy

Please select all the answers that are aligned with your learning objectives.

52 responses

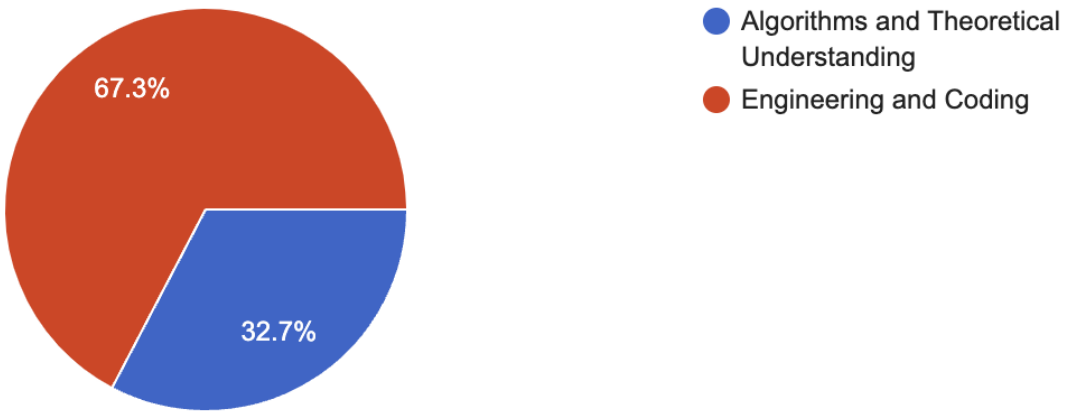


Survey Results: Preference

Preference: I prefer the course content to cover more of the following aspect

Copy

52 responses



Clarification

This is not the class if you want to

- learn programming
- learn basic machine learning
- learn how to use PyTorch

Prerequisites

Some requirements

- Proficiency in Python
- Basic Calculus and Linear Algebra
- Basic Probability and Statistics
- Foundations of Machine Learning

Topics

1. Explain the fundamental NLP techniques
   - Text classification
   - Language modeling
   - Word embeddings
   - Sequence labeling
   - Machine translation

Topics

1. Explain the fundamental NLP techniques
   - Text classification
   - Language modeling
   - Word embeddings
   - Sequence labeling
   - Machine translation
2. Advanced topics
   - Large language models
   - Text generation and summarization
   - In-context learning
   - NLP applications

Topics

1. Explain the fundamental NLP techniques
   - Text classification
   - Language modeling
   - Word embeddings
   - Sequence labeling
   - Machine translation
2. Advanced topics
   - Large language models
   - Text generation and summarization
   - In-context learning
   - NLP applications
3. Opportunities of working on some NLP problems
   - Final project

Schedule

Schedule

Dates	Topics	Slides	Note
Aug. 27, Aug. 29	Introduction, Text classification I		
Sep. 3, Sep. 5	No class		
Sep. 10, Sep. 12	Text classification II		**Sep. 12: Homework 1 out**
Sep. 17, Sep. 19	Word embeddings: LSA		
Sep. 24, Sep. 26	Word embeddings: Word2vec		**Sep. 25: Homework 2 out**
Oct. 1, Oct. 3	Language modeling: n-gram LMs		
Oct. 8, Oct. 10	Language modeling: Neural LMs		**Oct. 5: Project proposal due**
Oct. 17	Statistical machine translation		
Oct. 22, Oct. 24	Sequence-to-sequence models		**Oct. 18: Homework 3 out**
Oct. 29, Oct. 31	Text generation and summarization		
Nov. 5, Nov. 7	Contextualized word embeddings		**Nov. 4: Mid-term report due**
Nov. 12, Nov. 14	Transformers, GPT, and BERT		
Nov. 19, Nov. 21	Efficient fine-tuning		**Nov. 13: Homework 4 out**
Nov. 26	In-context Learning and Chain-of-thoughts		
Dec. 3, Dec. 5	NLP applications		**Dec. 5: Final project due (tentative)**

On course webpage

Assignments

- No exam

Assignments

- No exam
- Four homeworks
  - $15\% \times 4 = 60\%$

Assignments

- No exam
- Four homeworks
  - $15\% \times 4 = 60\%$
- One final project (40%)

Assignments

- No exam
- Four homeworks
  - $15\% \times 4 = 60\%$
- One final project (40%)
  - Proposal: 10%
  - Mid-term report: 10%
  - Final presentation: 10%
  - Final project report: 10%
  - 2 – 3 students per group

Policy: Late penalty

Homework submission will be accepted up to 72 hours late, with 20% deduction per 24 hours on the points as a penalty.

For example,

- Deadline: Sept. 15th, 11:59 PM
- Submission timestamp: Sept. 17th, 9:00 AM ($\leq$ 48 hours)
- Original points of a homework: 10
- Actual points:

$$10 \times (1 - 40\%) = 6.0 \tag{1}$$

It is usually better if students just turn in what they have in time.

Policy: Late Penalty (II)

- It's the students' responsibility to double check their submission and make sure you submit the correct file.
- If a student submits one homework via multiple files/times, we will use the latest timestamp for deciding and calculating the late penalty.
- Due to possible internet traffic issue, we can waive the late submission within 15 minutes after the deadline.

Textbooks

No single textbook is perfectly aligned with the course content.

- Textbook
  - Eisenstein, *Natural Language Processing*, 2018

All free online

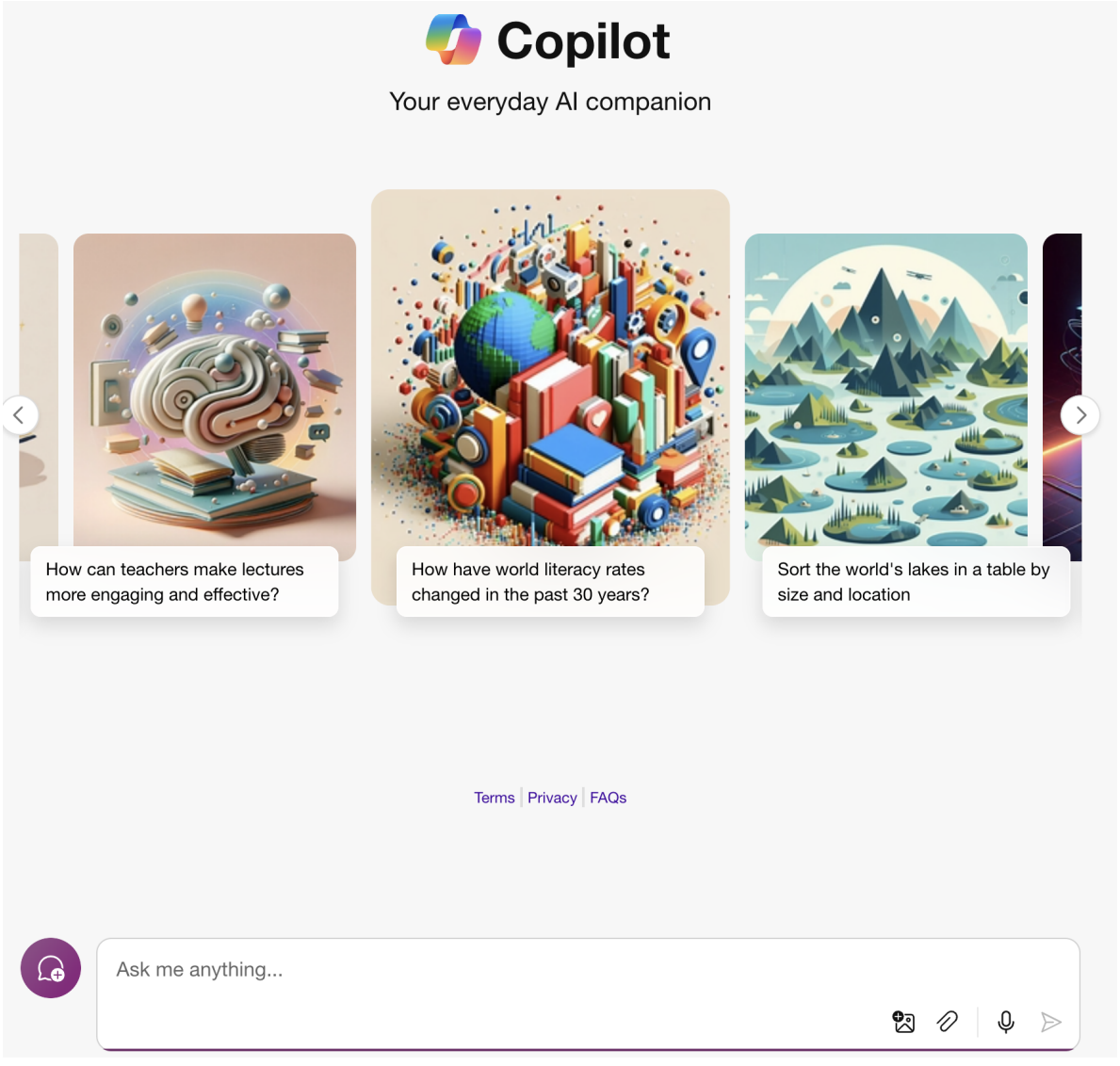
Textbooks

No single textbook is perfectly aligned with the course content.

- Textbook
  - Eisenstein, *Natural Language Processing*, 2018
- Additional textbooks
  - Jurafsky and Martin, *Speech and Language Processing*, 3rd Edition, 2020
  - Shalev-Shwartz and Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, 2014
  - Goodfellow, Bengio and Courville, *Deep Learning*, 2016

All free online

NLP Application Example: Microsoft Copilot



Free access via your UVA account

NLP Application Example: Grammarly

Great Writing, Simplified

Compose bold, clear, mistake-free writing with Grammarly's AI-powered writing assistant.

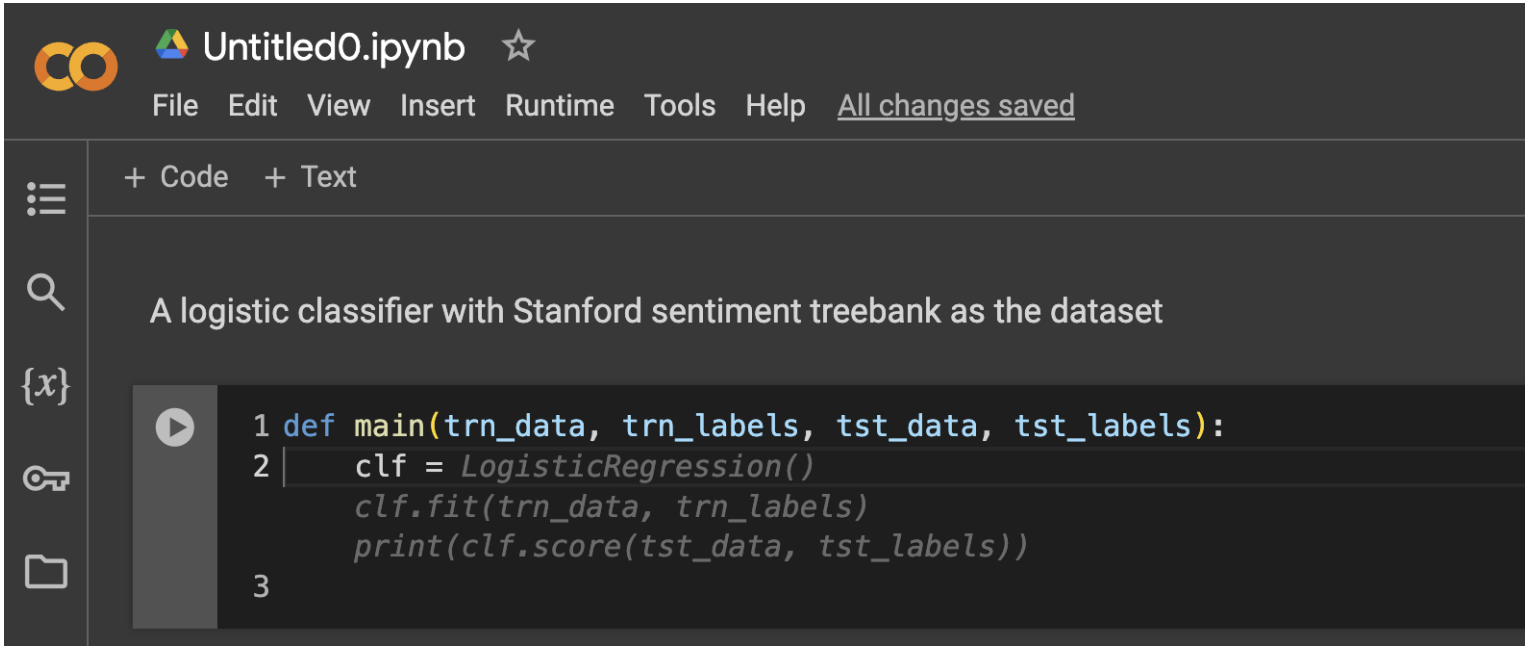
Add to Safari It's free

20 million people use Grammarly to improve their writing

https://www.grammarly.com/

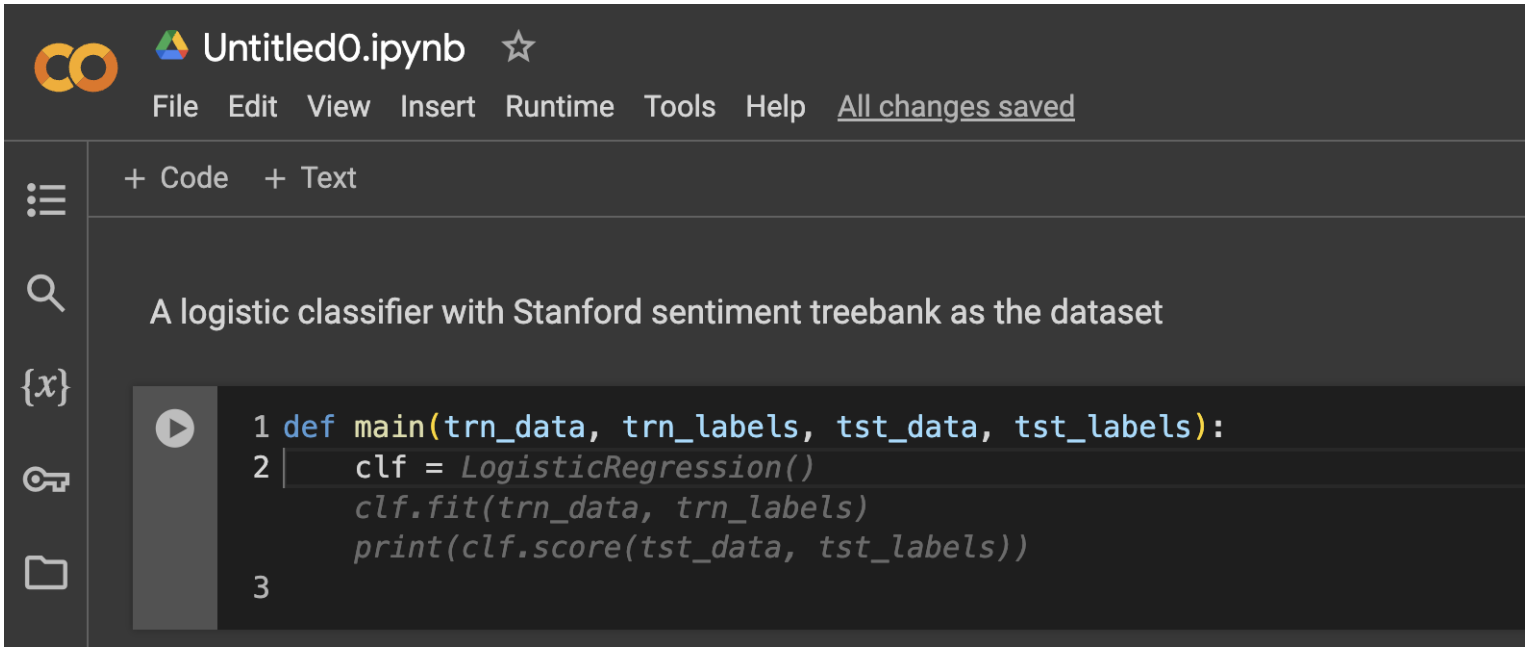
NLP Application Example: Code Generation

On Google Colab



NLP Application Example: Code Generation

On Google Colab



We encourage to use Google Colab for our homework assignments and the final project.

Policy: Collaboration and Grading

- About using Generative AI
  - Students are allowed to use generative AI for assignments and projects
  - Students will be responsible for the content and the correctness of the submitted answers
- For more collaboration policy, please refer to the course webpage

Generative AI Defending Classical Methods

You

Is statistical natural language processing still useful?

Copilot

Yes, statistical natural language processing (SNLP) is still useful, though its role has evolved with the advent of deep learning techniques. Here are a few reasons why SNLP remains relevant:

1. **Simplicity and Efficiency**: SNLP methods are often simpler and more computationally efficient compared to deep learning models. This makes them suitable for applications with limited computational resources.
2. **Interpretability**: Statistical models are generally more interpretable than complex neural networks. This can be crucial in applications where understanding the decision-making process is important.
3. **Baseline Models**: SNLP methods serve as strong baselines for evaluating more complex models. They provide a benchmark to measure the performance improvements offered by advanced techniques.
4. **Hybrid Approaches**: Combining statistical methods with neural networks can enhance performance. For example, statistical methods can be used for feature extraction, which can then be fed into neural network models.
5. **Specific Use Cases**: In some specific tasks, such as certain types of text classification or information retrieval, SNLP methods can still perform competitively [1] [2].

While deep learning has brought significant advancements in NLP, statistical methods continue to play a valuable role, especially in scenarios where simplicity, efficiency, and interpretability are prioritized.

Is there a particular aspect of NLP you're interested in exploring further?

Learn more 1 link.springer.com 2 hbr.org 3 cis.upenn.edu 4 en.wikipedia.org

The Design of This Course

- Topic-based discussion
  - E.g., text classification
- Each topic has examples from statistical NLP and neural NLP methods
  - E.g., logistic regression vs. feed-forward NN
- Cutting-edge NLP techniques will be covered
  - E.g., in-context learning, parameter-efficient tuning

Question?