# Natural Language Processing

## Course Administration & Presentation

Master Degree in Computer Engineering
University of Padua
Lecturer : Giorgio Satta

https://stem.elearning.unipd.it/course/view.php?id=7937

# Lecturer

**Giorgio Satta**
satta@dei.unipd.it
http://www.dei.unipd.it/~satta/

**Class hours**
Wed 12:30–14:30, **under email appointment**
Building: DEI/G, fourth floor
Zoom meeting: https://unipd.zoom.us/j/82934199797

# Schedule

| | **Mon** | **Tue** | **Wed** |
|---|---|---|---|
| 12:30–13:30 | | | class hours |
| 13:30–14:30 | | | class hours |
| 14:30–15:30 | | | |
| 15:30–16:30 | | | |
| 16:30–17:30 | room Ce | | room Ce |
| 17:30–18:30 | room Ce | | room Ce |

# Content outline

| LECTURE | WEEK | TOPIC | BOOK & CHAPTERS |
|---|---|---|---|
| | | **PRELIMINARIES** | |
| 1 | 1 | Natural language processing: Introduction | Introduction to Natural Language Processing, Eisenstein, Chapter 1 |
| 2 | 1 | Essentials of linguistics | Slides from the lecture |
| 3 | 2 | Text normalization | Speech and Language Processing, Jurafsky and Martin, Chapter 2 |
| | | **LARGE LANGUAGE MODELS** | |
| 4 | 2, 3 | Word embeddings | Speech and Language Processing, Jurafsky and Martin, Chapter 6<br>Introduction to Natural Language Processing, Eisenstein, Chapter 14 |
| 5 | 3, 4 | Language models | Speech and Language Processing, Jurafsky and Martin, Chapters 3, 7, |
| 6 | 4, 5 | Large language models | Speech and Language Processing, Jurafsky and Martin, Chapters 10, 11 |
| 7 | 5, 6 | ChatBots | Speech and Language Processing, Jurafsky and Martin, Chapter 12<br>Slides from the lecture |
| | | **STRUCURED PREDICTION** | |
| 8 | 6 | Part-of-Speech Tagging | Speech and Language Processing, Jurafsky and Martin, Chapter 17<br>Introduction to Natural Language Processing, Eisenstein, Chapter 7 |
| 9 | 7 | Phrase-Structure Parsing | Speech and Language Processing, Jurafsky and Martin, Chapter 18 |
| 10 | 7, 8 | Dependency Parsing | Speech and Language Processing, Jurafsky and Martin, Chapter 19 |
| 11 | 9 | Semantic Parsing | Speech and Language Processing, Jurafsky and Martin, Chapter 21 |
| | | **END-TO-END APPLICATIONS** | |
| 12 | 9, 10 | Machine Translation | Speech and Language Processing, Jurafsky and Martin, Chapter 13 |
| 13 | 11 | Question Answering | Speech and Language Processing, Jurafsky and Martin, Chapter 14<br>Introduction to Natural Language Processing, Eisenstein, Chapter 17 |
| 14 | 11, 12 | Dialogue Systems | Speech and Language Processing, Jurafsky and Martin, Chapter 15 |
| | | **CONCLUSIONS** | |
| 15 | 12 | Wrap-up | Slides from the lecture |

# Lab sessions

| SESSION | WEEK | TASK | MODEL |
|---|---|---|---|
| 1 | 4 | Word embedding | Skip-gram |
| 2 | 5 | Fine-tuning | BERT |
| 3 | 8 | Retrieval augmented generation | chatBot |

# Course requirements

Students should have **basic knowledge** of the following subjects

- calculus + linear algebra
- machine learning + deep learning
- probability theory + information theory
- computer algorithms + dynamic programming
- automata theory + rewriting grammars + formal languages
- Python + NumPy + PyTorch

The class also uses basic knowledge from **linguistics**: all of the working notions in linguistics will be properly introduced.

# Textbook

**Speech and Language Processing** (3rd ed., draft)
Dan Jurafsky and James H. Martin
January 12th, 2025
`https://web.stanford.edu/~jurafsky/slp3/`

Several chapters about machine learning, these will be given for granted.

# Textbook

Additional textbooks, for **consultation only**:

**Introduction to Natural Language Processing**
Jacob Eisenstein
MIT Press, October 2019

**Natural Language Processing**
**A Machine Learning Perspective**
Yue Zhang and Zhiyang Teng
Cambridge University Press, January 2021

# Additional resources

Additional material and resources available through **moodle page** of the class

- slides, videos and notebooks
- three forums for class, lab, and project discussion

# Project

Paper presentation:

- choose a scientific article from a given list of 2024 NLP conferences
- read, understand, and record a short video presentation

NLP system:

- for the announced task/dataset, provide
  - exploratory data analysis
  - baseline
  - neural approach
  - result comparison with state-of-the-art (SotA) and discussion
- report your project in a notebook

# Coursework

| | |
|---|---|
| Written test | 70% |
| Project | |
| 1. NLP system | 25% |
| 2. article presentation | 05% |
| | **100**% |

# Hard rules

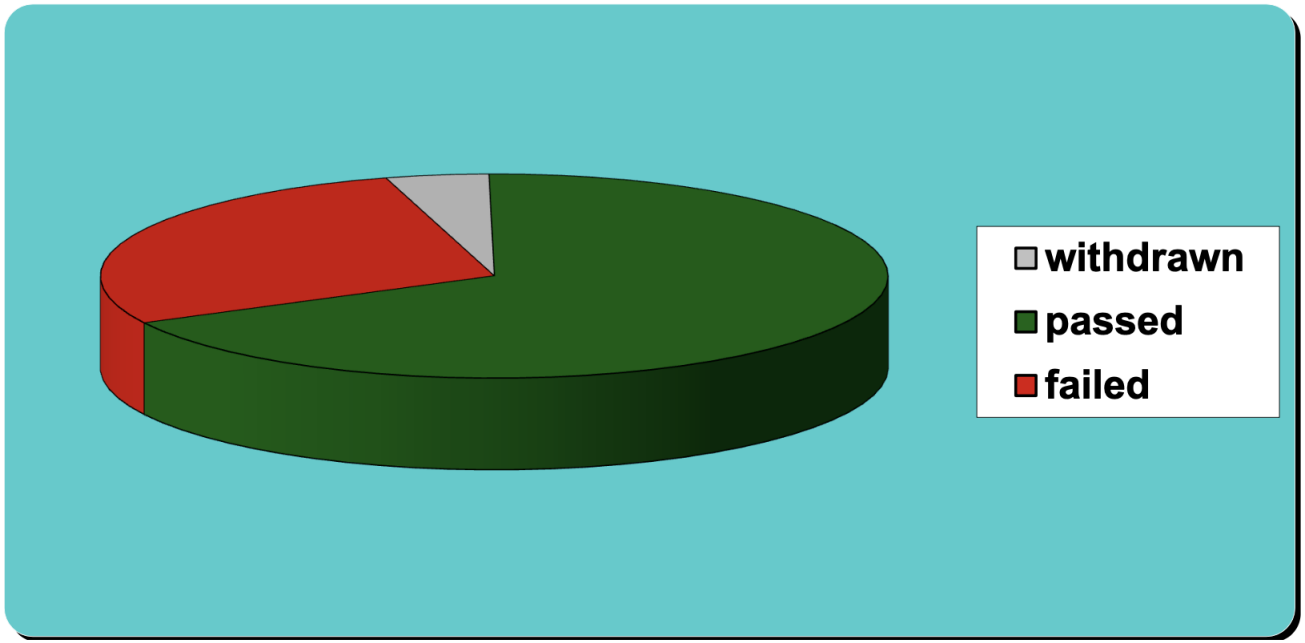
Project groups of **maximum** two people.

Project must be presented **before** earliest written test of the students in the group.

To pass the final exam you need to score 18/30 or above **for both** the written test and the project.

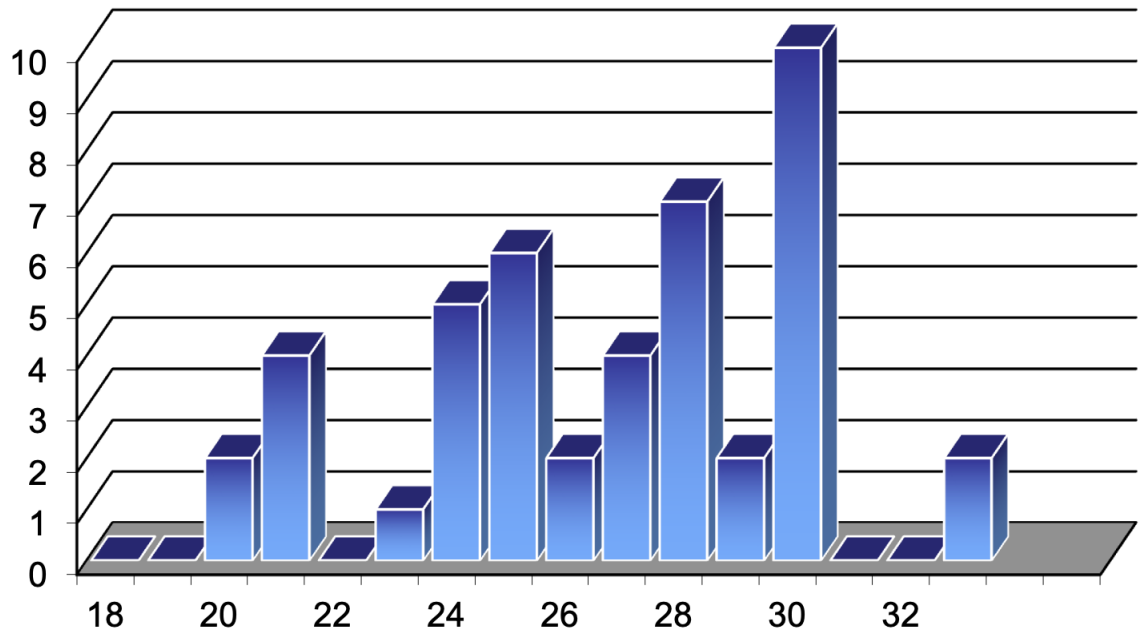
# Statistics

Academic year 2023-24



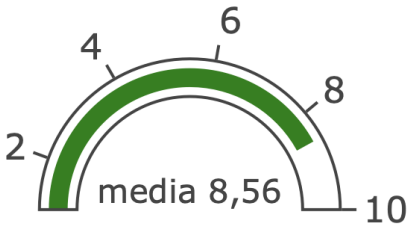
# Statistics

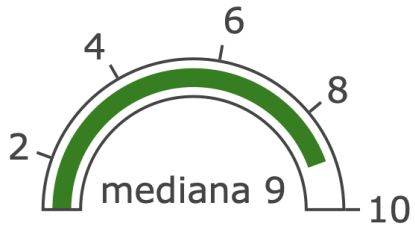
Academic year 2023-24



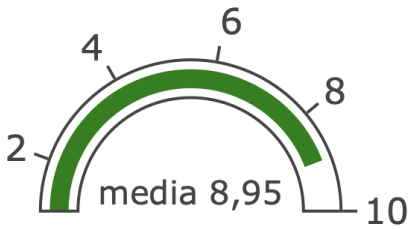
# Evaluation

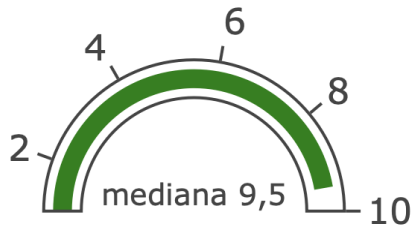
Academic year 2023-24



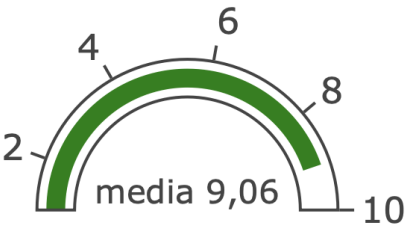



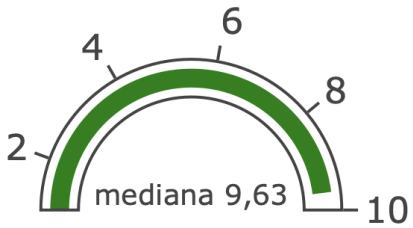

Satisfaction




Teaching




Course Organization