GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2025-2026

Introduction to Machine Learning and Feature Engineering for NLP (A005871)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0 **Study time 150 h**

Course offerings in academic year 2026-2027

A (semester 2)	Dutch	Gent
B (semester 1)	English	Gent

Lecturers in academic year 2026-2027

Hoste, Veronique	LW22	lecturer-in-charge
De Clercq, Orphée	LW22	co-lecturer
Lefever, Els	LW22	co-lecturer

Offered in the following programmes in 2026-2027	crdts	offering
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Language Technology)	5	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Language Technology)	5	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Language Technology)	5	A
Master of Arts in Advanced Studies in Linguistics (main subject Natural Language Processing: Theory and Practice)	6	A
Postgraduate Certificate Computer-Assisted Language Mediation	5	B
Preparatory Course Master of Arts in Multilingual Communication: a combination of at least two languages	5	A
Preparatory Course Master of Arts in Translation: a combination of at least two languages	5	A

Teaching languages

English, Dutch

Keywords

Data annotation, machine learning, feature-based systems

Position of the course

In this course, we focus on the architecture of data-driven machine learning (ML) systems. The complete workflow of a machine learning system is covered, ranging from data annotation to the training and evaluation of various feature-based machine learning architectures.
This course builds on the subject *NLP and Linguistic Analysis* and is taught over a period of 4 weeks (weeks 9–12).

Contents

The course deals with the following topics:

- Enriching data using various types of data annotation (expert annotation, games with a purpose, crowdsourcing).
- Evaluating the reliability of annotations using different metrics, such as Cohen's & Fleiss' Kappa and Krippendorff's Alpha.
- Automatically extracting various textual features (lexical, syntactic, semantic) from text and converting them into feature vectors, while also expanding these with information from external resources (word lists, corpora, treebanks, linguistic databases).
- The underlying algorithms and techniques of machine learning systems.
- Training and evaluating different systems based on various language technology applications (e.g., readability prediction, sentiment analysis).

- Ethical issues associated with data-driven machine learning.

Initial competences

Basic knowledge about NLP and Python.

Final competences

1 Students have basic knowledge of and insight into feature-based machine learning.
2 Students have the practical knowledge and skills to build machine-learning models, ranging from data annotation, over the construction of feature vectors, to the training and evaluation of machine learning systems.

Conditions for credit contract

This course unit cannot be taken via a credit contract

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Seminar, Lecture, Independent work

Study material

None

References

- Daniel Jurafsky & James Martin, "Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" (ed3).
- Mitchell, T. (1997). “Machine learning”. McGraw-hill New York.
- Python Software Foundation. Official Python documentation. http://www.python.org/doc/

Course content-related study coaching

Discussion forum on Ufora
Possibility to contact lecturers via e-mail

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral assessment

Examination methods in case of periodic assessment during the second examination period

Oral assessment

Examination methods in case of permanent assessment

Participation, Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

First session:
- Presence and participation in class (10%)
- Assignment (30%)
- Oral exam including a discussion on the assignment and assessment of theoretical knowledge (60%)

Second session:
- Assignment (40%)
- Oral exam(60%)

Calculation of the examination mark

A combination of continuous assessment (40%) and end-of-term evaluation (60%).
First exam session: To pass, the student must achieve an overall average of 10 out of 20 or higher, with a minimum score of 40% on both the continuous and end-of-term components. If the minimum score is not obtained for either component, the student can receive a maximum final score of 9 out of 20 for the course.
Second exam session: Both components must be retaken. To pass the course, the

student must again achieve an overall average of 10 out of 20 or higher, with a minimum score of 40% on both components. If the minimum score is not achieved for either component, the maximum final score remains 9 out of 20.

Facilities for Working Students

Class attendance is strongly recommended.
Limited possibility of feedback via e-mail, restricted to answering specific questions