



FIȘA UNITĂȚII DE CURS/MODULULUI

1. Date despre program

1.1. Instituția de învățământ superior	Universitatea de Stat din Moldova
1.2. Facultatea	Matematică și Informatică
1.3. Departamentul	Informatică
1.4. Domeniul de formare profesională	0613 Dezvoltarea produselor program și a aplicațiilor
1.5. Ciclul de studii	Studii superioare de master, ciclul II
1.6. Programul de studii	Inteligența Artificială și Știința Datelor

2. Date despre unitatea de curs/modul

2.1. Denumirea disciplinei	Procesarea Limbajului Natural						
2.2. Titularul activităților de curs	?						
2.3. Titularul activităților de laborator	?						
2.4. Anul de studii	2	2.5. Semestrul	1	2.6. Tipul de evaluare	E	2.7. Categoria formativă	SA

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1. Număr de ore pe săptămână	3	din care: 3.2 curs	2	3.3. laborator	1
3.4. Total ore în planul de învățământ	45	din care: 3.5 curs	30	3.6. laborator	15
<b>Distribuția fondului de timp pentru studiul individual:</b>					<b>ore</b>
Studiu după manual, suport de curs, bibliografie și notițe					40
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate					30
Pregătire laboratoare, teme, referate, portofolii					65
3.7. Total ore studiu individual	135				
3.8. Total ore pe semestru	180				
3.9. Număr de credite	6				

4. Precondiții de acces la unitatea de curs/modul

4.1. de curriculum	Baze de date, programare(Python), statistică pentru știința datelor, analiza și vizualizarea datelor, învățarea automată.
4.2. de competențe	Cunoștințe de programare, statistică și învățare automată.

5. Condiții de desfășurare

5.1. a cursului	Sală de curs cu tablă mare și proiector, laptop(cu aplicații instalate – biblioteci Python).
5.2. a laboratorului	Sală de laborator, calculatoare care au instalate Python.

6. Obiectivele disciplinei – rezultate așteptate ale învățării la formarea cărora contribuie parcurgerea și promovarea disciplinei

Cunoștințe	(C1) Dobândirea unor cunoștințe fundamentale a conceptelor și tehnicilor de bază utilizate în prelucrarea limbajului natural, inclusiv tokenizarea, lemmatizarea, analiza sintactică și semantică. (C2) Dobândirea unor cunoștințe aprofundate despre reprezentările vectoriale ale cuvintelor și documentelor (de exemplu, Word2Vec, TF-IDF) și modelele de limbaj avansate (de exemplu, N-gram, Transformeri). (C3) Dobândirea unor cunoștințe aprofundate a modului în care se aplică modele avansate, precum Rețelele
------------	---



	<p>Neuronale Recurente (RNN) și Mecanismele de Atenție, în sarcini complexe de NLP, cum ar fi traducerea automată și recunoașterea entităților cu nume (NER).</p> <p>(C4) Dobândirea unor cunoștințe aprofundate referitor la aplicațiile practice ale NLP în diverse domenii, cum ar fi analiza sentimentelor, clasificarea textului, modelarea topicurilor, și dezvoltarea de chatbot-uri.</p>
Abilități	<p>(A1) Abilitatea de a aplica tehnici de preprocesare a textului, cum ar fi tokenizarea, eliminarea cuvintelor oprite, și etichetarea părților de vorbire pentru a pregăti datele textuale pentru analiză.</p> <p>(A2) Abilitatea de dezvoltare și implementare a modelelor NLP pentru sarcini specifice, cum ar fi clasificarea textului, analiza sentimentelor, și recunoașterea entităților cu nume (NER).</p> <p>(A3) Abilitatea de a evalua și interpreta rezultatele obținute din modelele NLP, utilizând măsuri de performanță adecvate, cum ar fi acuratețea, precizia, și scorul F1.</p> <p>(A4) Abilitatea de a lucra cu unelte și biblioteci populare de NLP (de exemplu, NLTK, SpaCy, Hugging Face) pentru a dezvolta soluții eficiente și scalabile.</p> <p>(A5) Abilitatea de a dezvolta soluții NLP personalizate pentru aplicații specifice, cum ar fi analiza sentimentelor în social media sau traducerea automată.</p> <p>(A6) Abilități de vizualizare și prezentare a datelor textuale și a rezultatelor analitice, facilitând luarea deciziilor informate.</p>
Responsabilitate și autonomie	<p>(R1) Capacitatea de a proiecta, planifica și gestiona proiecte de NLP de la concept până la implementare, asumând responsabilitatea pentru rezultatele finale.</p> <p>(R2) Capacitatea de a lua decizii informate în ceea ce privește selecția și implementarea tehnicilor NLP adecvate pentru sarcini specifice, bazându-se pe o evaluare critică a cerințelor proiectului.</p> <p>(R3) Capacitatea de a continua învățarea și adaptarea la noile tehnologii și tendințe în NLP, menținându-se la curent cu evoluțiile din domeniu.</p> <p>(R4) Capacitatea de a colabora eficient cu colegii și echipele multidisciplinare în proiecte de NLP, contribuind cu expertiza personală și respectând contribuțiile altora.</p> <p>(R4) Capacitatea de a identifica problemele complexe în domeniul NLP și de a găsi soluții inovatoare și eficiente, asumându-și responsabilitatea pentru soluționarea acestora.</p> <p>(R5) Dezvoltarea capacității de a analiza critic diferitele abordări și tehnici în NLP, evaluând avantajele și limitările fiecărei metode.</p>

## 7. Conținutul unității de curs/modulului

7.1. Curs	Numărul de ore
-----------	----------------



<i>C1. Introducere în Prelucrarea Limbajului Natural.</i> Prezentarea generală a NLP și aplicațiile sale. Context istoric și evoluția NLP. Provocări în înțelegerea limbajului uman.	2
<i>C2. Prelucrarea Preliminară a Textului.</i> Tokenizare, stemming și lematizare. Eliminarea cuvintelor oprite. Etichetarea părților de vorbire.	2
<i>C3. Modele în Spațiu Vectorial.</i> Bag of Words (BoW). Term Frequency-Inverse Document Frequency (TF-IDF). Reprezentări vectoriale ale cuvintelor: Word2Vec și GloVe.	2
<i>C4. Sintaxă și Analiză Gramaticală.</i> Structura propozițiilor și arbori sintactici. Analiză de dependență. Analiză de constituire.	2
<i>C5. Modele de Limbaj.</i> Modele N-gram. Introducere în modelele neuronale de limbaj. Modele secvență-la-secvență și transformatori.	2
<i>C6. Recunoașterea Entităților cu Nume (NER).</i> Introducere în NER. Abordări bazate pe reguli și învățare automată. Aplicații ale NER în extracția informațiilor.	2
<i>C7. Analiza Sentimentelor.</i> Înțelegerea analizei sentimentelor. Tehnici pentru analiza sentimentelor: bazate pe lexic, bazate pe învățare automată. Analiza sentimentelor în datele de pe rețelele sociale.	2
<i>C8. Modelarea Topicurilor.</i> Introducere în modelarea topicurilor. Alocarea Dirichlet Latentă (LDA). Aplicații ale modelării topicurilor în clasificarea documentelor.	2
<i>C9. Clasificarea Textului.</i> Prezentarea generală a clasificării textului. Naive Bayes, Mașini cu Vectori Suport (SVM) și abordări de învățare profundă. Aplicații reale: detectarea spamului, categorizarea știrilor.	2
<i>C10. Modelarea Secvențelor și Rețele Neuronale Recurente (RNN).</i> Bazele RNN. Rețele cu Memorie pe Termen Lung (LSTM). Cazuri de utilizare: generarea de text, traducerea automată.	2
<i>C11. Mecanisme de Atenție și Transformatori.</i> Introducere în mecanismele de atenție. Arhitectura modelului Transformer. Aplicații: modelele BERT și GPT.	2
<i>C12. Sisteme de Dialog și Chatbot-uri.</i> Proiectarea interfețelor conversaționale. Chatbot-uri bazate pe reguli și AI. Evaluarea performanței chatbot-urilor.	2
<i>C13. Traducerea Automată.</i> Istoria și abordările în traducerea automată. Modele de traducere automată neuronală. Metrici pentru evaluarea calității traducerii.	2
<i>C14. NLP în Contexte Multilingve.</i> Provocările NLP pentru limbile non-engleze. Modele NLP translinguale și multilingve. Studii de caz în aplicații NLP multilingve.	2
<i>C15. Considerații Etice și Direcții Viitoare în NLP.</i> Părtinire și echitate în modelele de NLP. Utilizarea etică a NLP în aplicații. Tendințe viitoare și domenii de cercetare în NLP.	2
<i>Total prelegeri:</i>	30



**Bibliografie:**

1. Jurafsky, D., & Martin, J. H. (2019). *Speech and Language Processing*. 3rd Edition. Prentice Hall.
2. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
3. Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Morgan & Claypool.
4. Collobert, R., et al. (2011). "Natural Language Processing (Almost) from Scratch." In *Journal of Machine Learning Research*, 12: 2493-2537.
5. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
6. Pang, B., & Lee, L. (2008). "Opinion Mining and Sentiment Analysis." In *Foundations and Trends in Information Retrieval*.
7. Vaswani, A., et al. (2017). "Attention is All You Need." In *Advances in Neural Information Processing Systems*.

<b>7.2. Laborator</b>		<b>Numărul de ore</b>
L1. Tokenizarea și analiza sintactică în Python.		3
L2. Analiza sentimentului în texte cu Python.		3
L3. Clasificarea textului.		3
L4. Explorarea și utilizarea modelelor Transformer, cum ar fi BERT și GPT.		3
L5. Dezvoltarea unui chatbot simplu.		3
<i>Total laborator:</i>		15

**Bibliografie:**

1. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
2. Pang, B., & Lee, L. (2008). "Opinion Mining and Sentiment Analysis." In *Foundations and Trends in Information Retrieval*.
3. Jurafsky, D., & Martin, J. H. (2019). *Speech and Language Processing*. 3rd Edition. Prentice Hall.
4. Sudharsan Ravichandiran (2021). *Getting Started with Google BERT*. Packt Publishing
5. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta & Harshit Surana (2020). *Practical Natural Language Processing*. O'Reilly Media
6. Jyotika Singh (2023). *Natural Language Processing in the Real World*. CRC Press
7. Rajesh Arumugam, Rajalingappaa Shanmugamani (2018). *Hands-On Natural Language Processing with Python*. Packt Publishing
8. Dipanjan Sarkar (2019). *Text Analytics with Python. A Practitioner's Guide to Natural Language Processing*. Apress

**8. Evaluare**

Nota semestrială			
Atestarea 1	Atestarea 2	Evaluare curentă	Lucrul individual
25%	25%	25%	25%

Nota finală	
Nota semestrială	Nota la examen
60%	40%

Standard minim de performanță
Standard minim (cunoștințe și aptitudini necesare pentru nota 5)
<ul style="list-style-type: none"> <li>• prezența la cursuri și seminarii conform cerințelor generale ale facultății</li> </ul>



## Universitatea de Stat din Moldova



- cunoașterea conceptelor de bază din NLP
- abilitatea de a efectua analiza sintactică în Python.
- capacitatea de a implementa analiza sentimentelor utilizând bibliotecile Python

Evaluarea în cadrul disciplinei „Procesarea Limbajului Natural” se realizează formativ - activitatea la orele de laborator, verificarea orală sau scrisă curentă, două lucrări de atestare (la mijloc și la sfârșit de semestru), și sumativ (final) – 4 lucrări independente, realizate pe parcursul semestrului și examenul final la disciplină. Nota generală la disciplină însumează nota de la examen și cea semestrială, în proporție de 40 și, respectiv, 60 la sută, și apreciază gradul de corespundere cu finalitățile scontate: cunoștințele și competențele acumulate, abilitatea de a aplica cunoștințele, gradul de integrare a cunoștințelor de către studenți etc.

Data completării

25.08.2024

Titular de disciplină

Dr., conf. univ. ?

Data avizării în departament

Șef departament