



# **CIC0269 - Processamento de Linguagem Natural**

## **Plano de Ensino 2022/1**

Prof. Dr. Vinícius Ruela Pereira Borges

[viniciusrpb@unb.br](mailto:viniciusrpb@unb.br)

<http://viniciusrpb.github.io/>

## **1 Dados gerais**

Pré-requisitos: CIC0090 - Estruturas de Dados

Carga horária: 2 horas teóricas, 2 horas práticas

Dia e hora: 3ª-feira e 5ª-feira, 16:00h às 17:50h

Linguagem de programação: Python  $\geq 3.5$  (APIs numpy, sklearn, nltk, keras-tensorflow, pytorch, fast.ai).

## **2 Objetivo**

O curso “Processamento de Linguagem Natural” tem como objetivo principal o estudo de estratégias e a construção de modelos inteligentes de maneira a viabilizar o processamento e o entendimento de línguas naturais pelos computadores modernos.

## **3 Ementa**

Introdução ao processamento de linguagem natural. Níveis de representação e processamento linguístico: fonética e fonologia, morfologia, sintaxe, semântica, discurso e pragmática. Construção e anotação de corpos de texto. Redes neurais profundas e modelos de linguagem. Tarefas de processamento de linguagem natural: classificação de textos, extração de informações e reconhecimento de entidades nomeadas.

## **4 Metodologia**

- Aulas teóricas: exposição do conteúdo teórico por parte do professor conforme o cronograma previsto de aulas;
- Aulas práticas: desenvolvimento e construção de modelos a partir dos métodos estudados nas aulas teóricas. Algumas tarefas no Kaggle podem ser passadas durante o horário da aula;

## 5 Sistema de Avaliação

As atividades avaliativas da disciplina Processamento de Linguagem Natural compreendem atividades em sala de aula na plataforma Kaggle<sup>1</sup> (T), e um projeto (P). Todas as atividades avaliativas descritas a seguir possuem valor máximo de 10,0 pontos.

### 5.1 Projeto (P)

No projeto, o(a) aluno(a) deverá escolher um tema, problema ou estudo de caso, elaborar um método (metodologia ou processo) baseado em aprendizado de máquina, mineração de dados ou processamento de linguagem natural, realizar experimentações para avaliação e validação e discutir os resultados. O projeto deverá ser apresentado no horário da aula em dia a ser divulgado. Os critérios de avaliação do projeto serão fornecidos em momento apropriado, juntamente com a especificação técnica do projeto.

O(a) aluno(a) que atrasar a entrega de alguma parte do projeto por motivos de saúde, deverá comunicar ao professor em até 24 horas após a data final de entrega do respectivo projeto. Projetos não serão recebidos ou corrigidos caso sejam entregues após 5 dias de atraso e receberão penalização de 10% da nota por dia de atraso.

### 5.2 Atividades

A nota das atividades  $T$  é calculada pela média aritmética das  $M$  atividades enviadas à plataforma Kaggle e tarefas no Aprender3/Moodle, como mostra a Eq. (1):

$$T = \frac{T_1 + \dots + T_M}{M}. \quad (1)$$

### 5.3 Menção Final

Nessa modalidade remota, a média final MF compreende as notas  $P$  e  $T$ , MF será calculada para  $P \geq 5,0$  e  $T \geq 5,0$  como mostra a Eq. 2:

$$MF = \frac{7,0 \times P + 3,0 \times T}{10,0}, \quad (2)$$

caso contrário,  $MF$  será calculada como:

$$MF = \min(P, T) \quad (3)$$

Conforme o regimento da Universidade de Brasília, a menção final do(a) aluno(a) será determinada associando-se  $NF$  de acordo com os critérios abaixo:

Menção final	MF
SS (Superior)	$9,0 \leq MF \leq 10,0$
MS (Médio Superior)	$7,0 \leq MF < 9,0$
MM (Médio)	$5,0 \leq MF < 7,0$
MI (Médio Inferior)	$3,0 \leq MF < 5,0$
II (Inferior)	$0 \leq MF < 3,0$

O(a) aluno(a) que não obtiver frequência mínima de 75% em relação ao número total de aulas estará reprovado(a) por faltas, recebendo menção final SR, independentemente do valor da Nota Final  $NF$ .

---

<sup>1</sup><http://www.kaggle.com>

## 6 Plataforma Aprender3/Moodle

A plataforma educacional Aprender3<sup>2</sup> será utilizada para apoiar a disciplina, isto é, para disponibilização do material didático, especificações de atividades avaliativas, informes, comunicados e notas. A comunicação entre o professor e os alunos ocorrerá **oficialmente** pelo Fórum de Avisos do Aprender3.

Para acesso ao Aprender3/Moodle da turma, o(a) aluno(a) deve se cadastrar como usuário (basta preencher o formulário de cadastramento na página <http://aprender3.unb.br>), ou caso já esteja cadastrado, o(a) aluno(a) deve se inscrever<sup>3</sup> utilizando a senha:

## 7 Atendimento

O atendimento ocorrerá exclusivamente com o professor - não temos monitores, pois é a primeira vez que a disciplina é ofertada. O professor divulgará no primeiro dia aula o dia e horário para atendimento presencial. A qualquer momento, dúvidas podem ser resolvidas enviando sua mensagem para [viniciusrpb@unb.br](mailto:viniciusrpb@unb.br).

## 8 Bibliografia

- Mitchell T. M., Machine Learning, McGraw-Hill Science/Engineering/Math, 414 páginas, 1997.
- Tan P., Steinbach M., Kumar V., Introduction to data mining, Pearson Education India, 769 páginas, 2005.
- Jurafsky, D. and Martin, J.H., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall, 2008.
- Goodfellow I., Bengio Y. and Courville A., Deep Learning, The MIT Press, 800 páginas, 2016.
- Bengfort B., Ojeda T., Bilbro R., Applied Text Analysis with Python, O'Reilly Media, 332 páginas, 2018.
- Haykin S. S., Neural networks and learning machines. 3rd ed., Prentice Hall, 2009.

---

<sup>2</sup><https://aprender3.unb.br/course/view.php?id=xxxx>

<sup>3</sup>A página da disciplina no Aprender3 receberá novas inscrições a partir de 03/06/2022.