

# TI-2400

# Processamento de

# Linguagem Natural

---

ANTONIO CARVALHO - TREINAMENTOS

# Bem vindos ao 6º semestre

---

Parabéns, vocês são Fatecanos, já enfrentaram e enfrentarão muitos desafios e dificuldades, porém antes de se abaterem lembrem-se que já venceram muitas batalhas e que o esforço e a persistência derrubam qualquer muralha.

# Disciplina

---

TI-2400 Processamento de Linguagem Natural

Objetivo :

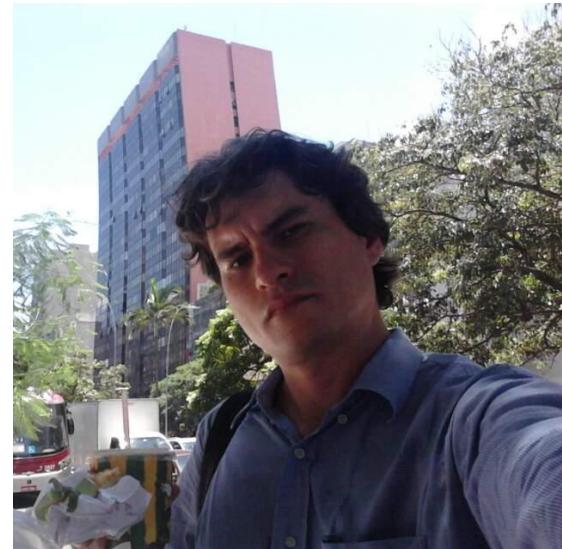
- Empregar técnicas e teorias de Inteligência Artificial aplicada ao Processamento de Linguagem Natural.
- Conhecer e aplicar técnicas e métodos para construção de recursos, ferramentas e aplicações de PLN.
- Construir mecanismos artificiais que permitam o entendimento de linguagem natural para realizar tarefas ou aplicações.
- Empregar linguagens de programação para o desenvolvimento de soluções.

Carga horária : 80 horas

# Professor

---

Antonio Rodrigues Carvalho Neto  
[antonio.carvalho@fatec.sp.gov.br](mailto:antonio.carvalho@fatec.sp.gov.br)



Ao enviar e-mails favor colocar no cabeçalho:  
FATEC - <TURMA>-<RA>-<NOME>-<Assunto>

# Ementa

---

# Ementa

---

- Processamento de linguagem natural.
- Aplicações de processamento de linguagem natural.
- Processamento de texto.
- Extração de características (*features*).
- Análise sintática.
- Interpretação semântica.
- Gramáticas.
- Descoberta de conhecimento em textos (*Knowledge Discovery in Texts*).
- Análise de sentimentos.
- Aprendizagem de máquina aplicada ao processamento de linguagem natural.

# Porque estamos aprendendo PLN com Python ?

Feb 2026	Feb 2025	Change	Programming Language	Ratings	Change
1	1		 Python	21.81%	-2.08%
2	4		 C	11.05%	+1.22%
3	2		 C++	8.55%	-2.82%
4	3		 Java	8.12%	-2.54%
5	5		 C#	6.83%	+2.71%
6	6		 JavaScript	2.92%	-0.85%
7	10		 Visual Basic	2.85%	+0.81%
8	15		 R	2.19%	+1.14%
9	7		 SQL	1.93%	-0.93%
10	9		 Delphi/Object Pascal	1.88%	-0.29%

Indice Tiobe 2026-fevereiro  
<https://www.tiobe.com/tiobe-index/>

# Competências

---

Desenvolver aplicativos para análise de sentimento,  
classificador de SPAM e *chatbots*

Utilizar as bibliotecas mais comuns para python  
relacionadas a *machine learning*

# Competências

---

Os níveis das competências serão divididos em 4 níveis:

- Conceitual
- Prática
- Domínio do conhecimento
- Domínio da competência

Os exercícios estão sendo reorganizados para contemplar estas competências.

Semana	Conteúdo Programático	Objetivo	Metodologia
1	<ul style="list-style-type: none"> <li>• <i>Introdução à PLN</i></li> </ul>	<i>Mostrar o conteúdo programático e as avaliações a serem feitas na disciplina</i>	<i>Aula Expositiva, Demonstração</i>
2	<ul style="list-style-type: none"> <li>• <i>O que é PLN ? Principais aplicações no mercado</i></li> <li>• <i>Termos e conceitos utilizados</i></li> <li>• <i>Tecnologias utilizadas</i></li> </ul>	<i>Contextualizar o uso de PLN</i> <i>Mostrar os termos e tecnologias que serão utilizados no processo</i> <i>Introdução ao Jupiter</i>	<i>Aula Expositiva, Demonstração</i>
3	<ul style="list-style-type: none"> <li>• <i>Tokenização</i></li> <li>• <i>Regex</i></li> </ul>	<i>Introduzir a tarefa de dividir os textos em palavras por meio de iteração em listas de Strings</i> <i>Funcionamento e uso do Regex para pesquisa e manipulação de Strings</i> <i>Criação de um chatbot com Regex</i>	<i>Aula Expositiva, Demonstração</i>
4	<ul style="list-style-type: none"> <li>• <i>Documento</i></li> <li>• <i>Corpus</i></li> <li>• <i>Vocabulário</i></li> </ul>	<i>Estratificação de documentos de um corpus para criação do vocabulário</i> <i>Criação do vocabulário com base no texto</i> <i>Uso de bi-gram e n-gram</i>	<i>Aula Expositiva, Demonstração</i>
5	<ul style="list-style-type: none"> <li>• <i>Stop words</i></li> <li>• <i>Bag of words</i></li> <li>• <i>One Hot Vector</i></li> <li>• <i>TF e IDF</i></li> </ul>	<i>Definição da lista de stop words</i> <i>Organização das palavras em estruturas numéricas para comparação</i> <i>Term Frequency e Inverse Document Frequency</i>	<i>Aula Expositiva, Demonstração</i>
6	<ul style="list-style-type: none"> <li>• <i>Numpy e Matplotlib</i></li> </ul>	<i>Biblioteca utilizada na criação de arrays e vetores em python</i> <i>Introdução ao Matplotlib</i>	<i>Aula Expositiva, Demonstração</i>
7	<ul style="list-style-type: none"> <li>• <i>Pandas</i></li> </ul>	<i>Biblioteca para trabalhar com dataframes</i> <i>Criar Dataframes, Importação de CSV, criação de colunas e filtros</i>	<i>Aula Expositiva, Demonstração</i>
8	<ul style="list-style-type: none"> <li>• <i>Steamming</i></li> <li>• <i>Lemmatização</i></li> </ul>	<i>Extração dos steamings por meio da biblioteca NLTK</i> <i>Lemmatization utilizando a biblioteca Spacy</i>	<i>Aula Expositiva, Demonstração</i>
9	<ul style="list-style-type: none"> <li>• <i>Revisão de machine learning</i></li> <li>• <i>Funções de ativação</i></li> </ul>	<i>Revisão de Machine Learning</i>	<i>Aula Expositiva, Demonstração</i>
10	<ul style="list-style-type: none"> <li>• <i>Implementar o SGD em Python</i></li> </ul>	<i>Implementar um algoritmo de descida estocástica do gradiente em Python</i>	<i>Aula Expositiva, Demonstração</i>

Semana	Conteúdo Programático	Objetivo	Metodologia
11	• <i>AV1</i>	<i>Avaliação 1</i>	<i>Avaliação</i>
12	• <i>Classificação</i> • <i>Analizador de sentimento</i>	<i>Uso de biblioteca para criar sistema de classificação (Pytorch)</i> <i>Criação de um analisador de sentimento usando <b>one hot vector</b> e método de classificação linear como regressão logística</i>	<i>Aula Expositiva, Demonstração</i>
13	• <i>Categorização com Softmax</i>	<i>Categorizador de textos utilizando Softmax</i>	<i>Aula Expositiva, Demonstração</i>
14	• <i>Embeddings</i>	<i>Melhoria na análise de textos com o uso de Embeddings usando word2Vec</i>	<i>Aula Expositiva, Demonstração</i>
15	• <i>Analise de semântica</i> • <i>Named Entity Recognition</i>	<i>Localização de entidades importantes no texto</i>	<i>Aula Expositiva, Demonstração</i>
16	• <i>Recurring Neural Networks</i> • <i>Implementar software para identificar a próxima palavra do texto</i>	<i>O que são RNNs e qual a vantagem delas no processo de classificação LSTM e Seq2Seq</i>	<i>Aula Expositiva, Demonstração</i>
17	• <i>Chatbot</i>	<i>Criação de um chatbot usando LSTM</i>	<i>Aula Expositiva, Demonstração</i>
18	• <i>AV2</i>	<i>Avaliação 2</i>	<i>Avaliação</i>
19	• <i>Revisão de conceitos e Entrega de Notas</i> • <i>Deploy do chatbot</i>	<i>Revisão de conceitos e Entrega de Notas.</i> <i>Disponibilização de um servidor de chatbot e integração com o sistema RASA</i>	<i>Aula Expositiva, Demonstração</i>
20	• <i>AV3</i>	<i>AV3 - Prova envolvendo todos os conceitos ensinados ao longo do semestre, segundo o planejamento de aula</i> <i>Duração da prova de aproximadamente 2 horas, após este período as provas serão corrigidas e as notas entregue para os alunos.</i>	<i>Avaliação</i>

# Avaliação

---

# Critérios de Avaliação

---

- Prova AV1 – (Exercícios 30% - Prova 70% - Remoto)
- Prova AV2 – (Exercícios 30% - Prova 70% - Remoto)
- Prova Substitutiva AV3 – (Prova 100%- Presencial)

Calculo da média será a melhor nota entre

- $(AV1 + AV2) / 2$
- $(AV2 + AV3) / 2$
- $(AV1 + AV3) / 2$

# Plágio é crime

---

Quero lembrar a todos que o plágio é uma prática ilegal e antiética.

Os códigos submetidos serão rigorosamente verificados quanto à originalidade.

Caso sejam encontrados códigos semelhantes ou idênticos aos de outras fontes ou colegas, as submissões serão automaticamente desconsideradas e receberão nota zero (tanto o original quanto a cópia).

Por favor, respeitem as normas de honestidade acadêmica e garantam que todos os trabalhos sejam fruto do seu próprio esforço.

Em caso de pedido de revisão de nota, é preciso estar ciente que nota pode ser revista para cima ou para baixo

# Regras sobre as Avaliações

---

- Nos dias da avaliação presencial o aluno pode trazer apenas 1 folha de tamanho A4 com seu RA, NOME e TURMA com suas anotações na frente e no verso, as anotações devem estar escritas a mão e devem ser original (não pode ser cópia), não pode haver outros papéis colados na folha A4.
- Todas as submissões (presenciais e remotas) serão verificadas, se houver mais que 20% de plágio no relatório, serão descontados 1,0 da nota para cada 5% de plágio adicional detectado.
- Exemplo:
  - Trabalho entregue e avaliado com 9,0 porém relatório indicando 80% de plágio
  - $80\% - 20\% = 60\%$
  - $60\% / 5 = 12,0$  (pontos)
  - Nota Final  $9,0 - 12,0 = (-3,0)$  Nota Final

# Estrutura das aulas

---

- Aulas baseadas em hands-on com cases que envolvam o conteúdo da disciplina e as necessidades de projeto.
- Exemplos desenvolvidos serão disponibilizados no GitHub.
  - [www.github.com/antoniorcn/fatec-2026-1s.git](https://www.github.com/antoniorcn/fatec-2026-1s.git)
- O material de apoio estará disponível no Teams
- Apoio aos projetos

# Uso do celular e dos computadores

---

Peço que, durante o período de aula, os celulares e computadores sejam utilizados exclusivamente para atividades relacionadas ao conteúdo que está sendo ensinado.

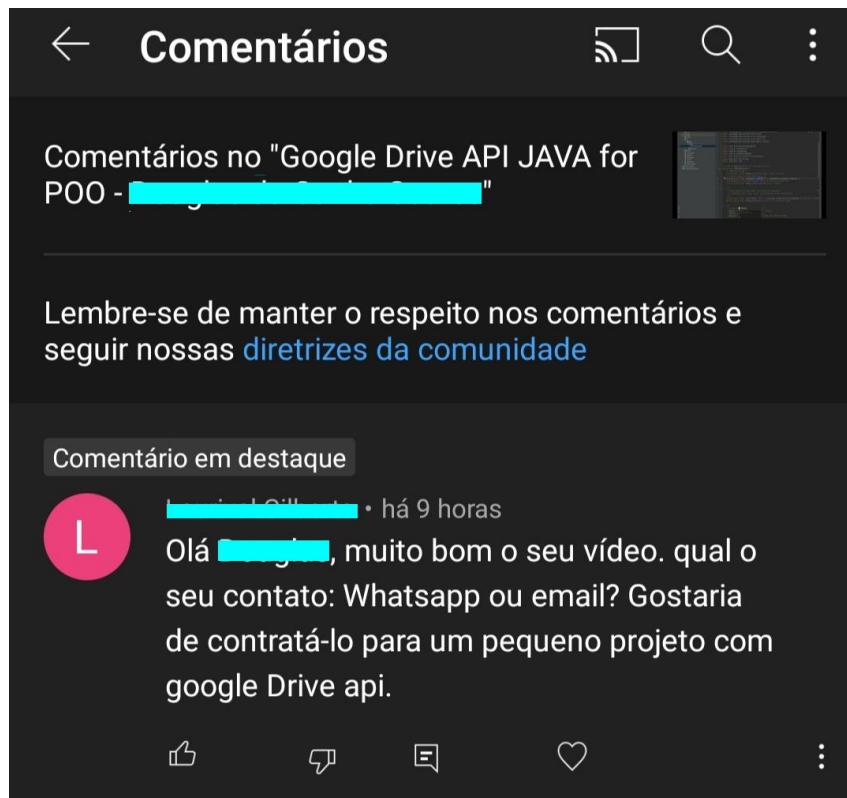
O acesso a redes sociais, jogos, vídeos e outros assuntos que não estão diretamente relacionados ao aprendizado pode comprometer a concentração e desempenho acadêmico sua e dos colegas, e não será tolerado.

Se precisar falar ao celular por voz ou texto peço que por gentileza se retire da sala de aula

# Sempre faça um bom trabalho

---

Façam um bom trabalho, o mercado está de olho em vocês.



# Outras informações

---

# Recursos Utilizados

---

**Anaconda** para baixar acesse o site →  
<https://www.anaconda.com/download#downloads>

Ou baixar tudo na mão

Para baixar o **Python** acesse o site → <http://www.python.org/download/> procure a versão 3.13 do Python

# Regras importantes

---

Ouça com atenção para não perder o conteúdo das aulas

Durante as aulas não utilizem as máquinas ou celulares para acessar redes sociais, notícias, youtube, email, e outros conteúdos não relativos à aula

Respeitem tudo e todos, principalmente os funcionários, colegas e professores

Cuidem do patrimônio, seja seu, do seu colega ou da faculdade

Ao deixarem a sala de aula/laboratório, verifiquem se:

- Janelas estão fechadas
- Máquinas estão desligadas (Computadores, Monitores, Datashow e TV)
- Carteiras / Cadeiras arrumadas
- Quadro apagado
- Ventiladores e Ar Condicionados desligados
- Lixo recolhido

No campus:

- Jogue o lixo no lixo
- Não faça barulho desnecessariamente, não converse alto

# Bibliografia

## BIBLIOGRAFIA BÁSICA

**FERREIRA, M., LOPES, M.** Para conhecer linguística computacional. São Paulo: Contexto, 2019

**RAJ, S.** Construindo Chatbots com Python: Usando Natural Language Processing e Machine Learning. São Paulo: Novatec, 2019.

**RUSSEL, M.A.** Mineração de Dados da Web Social: Análise de dados do Facebook, Twitter, Linkedin e outros sites de mídia social. São Paulo: Novatec, 2019.

## BIBLIOGRAFIA COMPLEMENTAR

**GÉRON, A.** Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow. Rio de Janeiro: Alta Books, 2019.

**HOBSON, L., COLE, H., HANNES, H.** Natural Language Processing in Action: Understanding, analyzing, and generating text with Python. Manning, 2019.

**MCKINNEY, W.** Python Para Análise de Dados: Tratamento de Dados com Pandas, NumPy e IPython. São Paulo: Novatec, 2018.

**MUELLER, J. P., MASSARON, L.** Python Para Data Science Para Leigos. Rio de Janeiro: Alta Books, 2020.

**TAKAHASHI, S., INOUE, I.** Guia Mangá de Análise de Regressão. São Paulo: Novatec, 2019.

**TAKAHASHI, S., INOUE, I.** Guia Mangá de Estatística. São Paulo: Novatec, 2010.

**VAJJATA, S., MAJUMDER, B., GUPTA, A., SURANA, H.** Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems. O'Reilly Media, 2020.

# Fica a Dica

---

Não presuma, tenha coragem de perguntar

Ainda que a dúvida pareça simplória, pergunte:  
você sempre sairá ganhando. (Livro: A Bagagem dos  
Viajantes)