

TI-2400

Processamento de Linguagem Natural

ANTONIO CARVALHO - TREINAMENTOS

A solid blue horizontal bar spanning the entire width of the slide at the bottom.

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

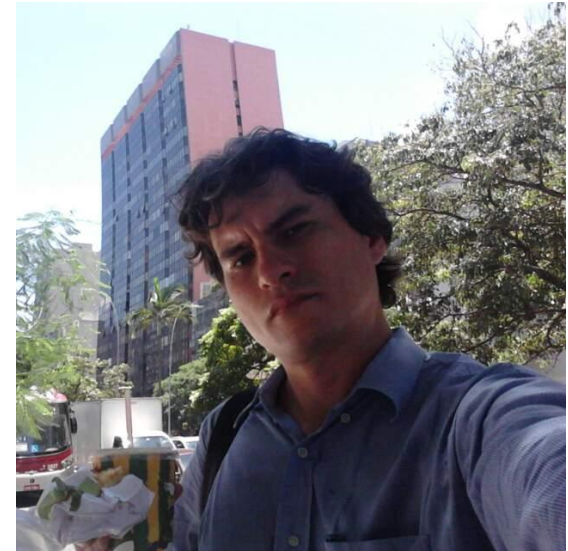
Objetivo :

- 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













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.

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

Semana	Conteúdo Programático	Objetivo	Metodologia
11	<ul style="list-style-type: none"> AV1 	Avaliação 1	Avaliação
12	<ul style="list-style-type: none"> Classificação Analizador de sentimento 	Uso de biblioteca para criar sistema de classificação (Pytorch) Criação de um analisador de sentimento usando one hot vector e método de classificação linear como regressão logística	Aula Expositiva, Demonstração
13	<ul style="list-style-type: none"> Categorização com Softmax 	Categorizador de textos utilizando Softmax	Aula Expositiva, Demonstração
14	<ul style="list-style-type: none"> Embeddings 	Melhoria na análise de textos com o uso de Embeddings usando word2Vec	Aula Expositiva, Demonstração
15	<ul style="list-style-type: none"> Análise de semântica Named Entity Recognition 	Localização de entidades importantes no texto	Aula Expositiva, Demonstração
16	<ul style="list-style-type: none"> Recurring Neural Networks Implementar software para identificar a próxima palavra do texto 	O que são RNNs e qual a vantagem delas no processo de classificação LSTM e Seq2Seq	Aula Expositiva, Demonstração
17	<ul style="list-style-type: none"> Chatbot 	Criação de um chatbot usando LSTM	Aula Expositiva, Demonstração
18	<ul style="list-style-type: none"> AV2 	Avaliação 2	Avaliação
19	<ul style="list-style-type: none"> Revisão de conceitos e Entrega de Notas Deploy do chatbot 	Revisão de conceitos e Entrega de Notas. Disponibilização de um servidor de chatbot e integração com o sistema RASA	Aula Expositiva, Demonstração
20	<ul style="list-style-type: none"> AV3 	AV3 - Prova envolvendo todos os conceitos ensinados ao longo do semestre, segundo o planejamento de aula Duração da prova de aproximadamente 2 horas, após este período as provas serão corrigidas e as notas entregue para os alunos.	Avaliação

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
- O material de apoio estará disponível no Teams
- Apoio aos projetos

Uso do celular e dos computadores

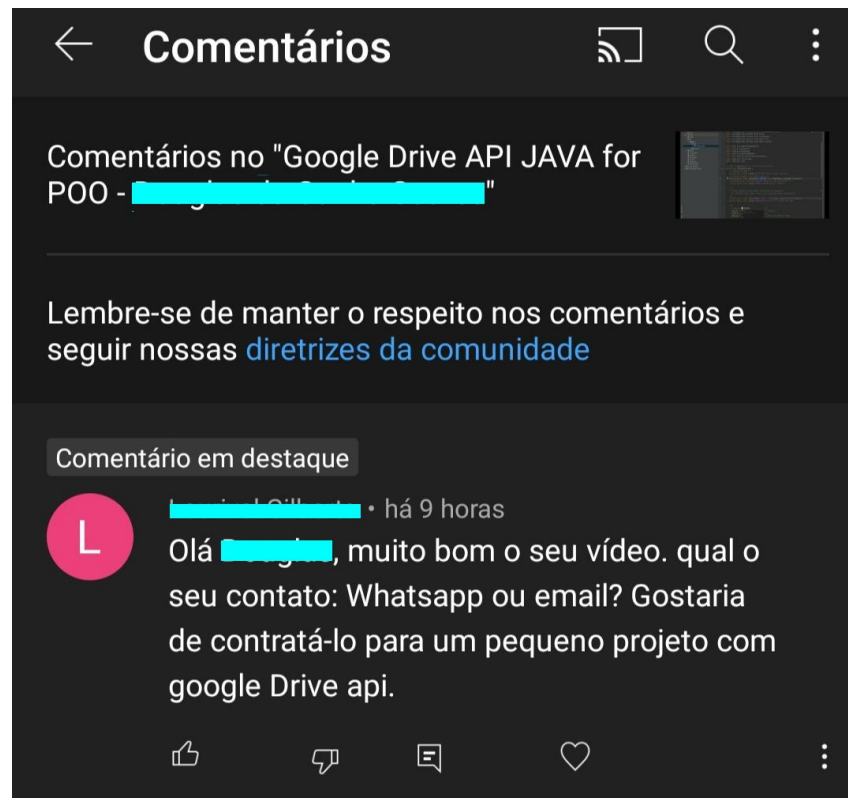
Peço que, durante o período de aula, os celulares e computadores sejam utilizados exclusivamente para atividades relacionadas ao conteúdo que está sendo ensinado.

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

Ouçá 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)