

PLANO DE ENSINO

IDENTIFICAÇÃO		
HABILITAÇÃO:	Pré-processamento de Dados	C. H. TOTAL: 80 horas
PROFESSOR (A):	Gerson do Nascimento Silva (PhD Student, UnB)	

PLANEJAMENTO			
BASES TECNOLÓGICAS			
<p>Ementa: Desenvolvimento de código-fonte (<i>script</i>) utilizando linguagem <i>python</i> para implementação, mostrando como:</p>			
<p>1. Uso da biblioteca “scikit-learn” (sklearn) e “pandas” :</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>a) Carregar dados de habitação, via sklearn (Boston);</p> <p>b) Criar dados simulados para regressão;</p> <p>c) Criar dados simulados para classificação;</p> <p>d) Criar dados simulados para armazenamento em <i>cluster</i>;</p> <p>e) Preparar um fluxo de trabalho de aprendizado de máquina;</p> <p>f) Converter recursos (características) categóricos em recursos numéricos;</p> <p>g) Imputar rótulos de classes ausentes;</p> <p>h) Imputar rótulos de classes ausentes usando método "vizinho próximo";</p> <p>i) Excluir instâncias com valores ausentes;</p> <p>j) Como fazer operações numéricas;</p> <p>k) Como encontrar <i>outliers</i>;</p> <p>l) Codificar recursos categóricos ordinais;</p> <p>m) Lidar com classes de desequilíbrio com redução da resolução;</p> <p>n) Como lidar com classes de desbalanceadas;</p> <p>o) Como lidar com <i>outliers</i>;</p> </td> <td style="width: 50%; vertical-align: top;"> <p>p) Imputar valores ausentes com médias;</p> <p>q) Codificação com vários rótulos;</p> <p>r) Codificação com recursos nominais categóricos;</p> <p>s) Processar recursos categóricos;</p> <p>t) Redimensionar recursos;</p> <p>u) Padronizar recursos;</p> <p>v) Padronizar dados "IRIS";</p> <p>w) Dividir dados <i>DateTime</i> ("features") para criar vários recursos;</p> <p>x) Calcular a diferença entre datas;</p> <p>y) Codificar os dias da semana;</p> <p>z) Tratar valores ausentes em uma série temporal;</p> <p>aa) Como introduzir o tempo "LAG" (lagged time-series), tempo de latência são muito usadas em análises econômicas;</p> <p>bb) Como lidar com "Janelas de Tempo";</p> <p>cc) Selecionar <i>DateTime</i> dentro de um intervalo;</p> <p>dd) Selecionar <i>DateTime</i> [formato (PM) ou (AM)] dentro de um intervalo</p> <p>ee) Como trabalhar itens em uma lista;</p> </td> </tr> </table>		<p>a) Carregar dados de habitação, via sklearn (Boston);</p> <p>b) Criar dados simulados para regressão;</p> <p>c) Criar dados simulados para classificação;</p> <p>d) Criar dados simulados para armazenamento em <i>cluster</i>;</p> <p>e) Preparar um fluxo de trabalho de aprendizado de máquina;</p> <p>f) Converter recursos (características) categóricos em recursos numéricos;</p> <p>g) Imputar rótulos de classes ausentes;</p> <p>h) Imputar rótulos de classes ausentes usando método "vizinho próximo";</p> <p>i) Excluir instâncias com valores ausentes;</p> <p>j) Como fazer operações numéricas;</p> <p>k) Como encontrar <i>outliers</i>;</p> <p>l) Codificar recursos categóricos ordinais;</p> <p>m) Lidar com classes de desequilíbrio com redução da resolução;</p> <p>n) Como lidar com classes de desbalanceadas;</p> <p>o) Como lidar com <i>outliers</i>;</p>	<p>p) Imputar valores ausentes com médias;</p> <p>q) Codificação com vários rótulos;</p> <p>r) Codificação com recursos nominais categóricos;</p> <p>s) Processar recursos categóricos;</p> <p>t) Redimensionar recursos;</p> <p>u) Padronizar recursos;</p> <p>v) Padronizar dados "IRIS";</p> <p>w) Dividir dados <i>DateTime</i> ("features") para criar vários recursos;</p> <p>x) Calcular a diferença entre datas;</p> <p>y) Codificar os dias da semana;</p> <p>z) Tratar valores ausentes em uma série temporal;</p> <p>aa) Como introduzir o tempo "LAG" (lagged time-series), tempo de latência são muito usadas em análises econômicas;</p> <p>bb) Como lidar com "Janelas de Tempo";</p> <p>cc) Selecionar <i>DateTime</i> dentro de um intervalo;</p> <p>dd) Selecionar <i>DateTime</i> [formato (PM) ou (AM)] dentro de um intervalo</p> <p>ee) Como trabalhar itens em uma lista;</p>
<p>a) Carregar dados de habitação, via sklearn (Boston);</p> <p>b) Criar dados simulados para regressão;</p> <p>c) Criar dados simulados para classificação;</p> <p>d) Criar dados simulados para armazenamento em <i>cluster</i>;</p> <p>e) Preparar um fluxo de trabalho de aprendizado de máquina;</p> <p>f) Converter recursos (características) categóricos em recursos numéricos;</p> <p>g) Imputar rótulos de classes ausentes;</p> <p>h) Imputar rótulos de classes ausentes usando método "vizinho próximo";</p> <p>i) Excluir instâncias com valores ausentes;</p> <p>j) Como fazer operações numéricas;</p> <p>k) Como encontrar <i>outliers</i>;</p> <p>l) Codificar recursos categóricos ordinais;</p> <p>m) Lidar com classes de desequilíbrio com redução da resolução;</p> <p>n) Como lidar com classes de desbalanceadas;</p> <p>o) Como lidar com <i>outliers</i>;</p>	<p>p) Imputar valores ausentes com médias;</p> <p>q) Codificação com vários rótulos;</p> <p>r) Codificação com recursos nominais categóricos;</p> <p>s) Processar recursos categóricos;</p> <p>t) Redimensionar recursos;</p> <p>u) Padronizar recursos;</p> <p>v) Padronizar dados "IRIS";</p> <p>w) Dividir dados <i>DateTime</i> ("features") para criar vários recursos;</p> <p>x) Calcular a diferença entre datas;</p> <p>y) Codificar os dias da semana;</p> <p>z) Tratar valores ausentes em uma série temporal;</p> <p>aa) Como introduzir o tempo "LAG" (lagged time-series), tempo de latência são muito usadas em análises econômicas;</p> <p>bb) Como lidar com "Janelas de Tempo";</p> <p>cc) Selecionar <i>DateTime</i> dentro de um intervalo;</p> <p>dd) Selecionar <i>DateTime</i> [formato (PM) ou (AM)] dentro de um intervalo</p> <p>ee) Como trabalhar itens em uma lista;</p>		
TEORIA	PRÁTICA		
<p>-video aula: (playlist do PhD Francisco A. Rodrigues – USP)</p> <ul style="list-style-type: none"> • “Estatísticas”; • “Probabilidades”; • “Redes Complexas”; • “Processos Estocásticos”; • “Ciência de Dados”; 	<p>- Pré-processamento: (representação e a qualidade dos dados)</p> <p>passo importante no processo de tratar dados. A frase "<i>garbage in, garbage out</i>" é particularmente aplicável aos projetos da trilha de IA (data mining, data science e machine learning). Os métodos de coleta de dados geralmente são pouco controlados, resultando em valores de intervalo <i>out-of-range</i> (por exemplo, renda: -100), combinações de dados impossíveis (por exemplo, sexo: masculino, grávidas: sim), <i>missing values</i>, etc. As análise de dados que não foram cuidadosamente selecionados para tais problemas pode produzir resultados enganosos. Assim, a representação e a qualidade dos dados são antes de tudo uma análise.</p>		

ANÁLISE DA REALIDADE
Pré-Requisito
<ol style="list-style-type: none"> 1. Lógica de Programação; 2. Programação Orientada a Objetos; 3. Análise de Sistemas Orientada a Objeto; 4. Fundamentos e Modelagem de banco de Dados.

Necessidade da turma:

- Programação intermediário em *python* ou outra linguagem de programação que traga subsídios para entendimento dos conceitos computacionais.

PROJEÇÃO DE FINALIDADES

Objetivos Gerais (Competências)

Compreender os conceitos de IA e seus recursos e capacidade para implementar código-fonte reproduzível.

Objetivos Específicos (Habilidades)

-Capacidade para implementar código-fonte (uso de *python* ou outras tecnologias com mesma finalidade);
-Capacidade para implementar APIs para o pré-processamento de dados;

-Ser capaz de implementar código-fonte reproduzível utilizando conceitos anteriores.

FORMAS DE MEDIAÇÃO

Procedimentos Metodológicos

-Elaborar modelos de implementação de código-fonte que valide as teorias, com uso de exemplos;
-Propor trabalhos práticos sobre o assunto;
-Corrigir os trabalhos práticos;
-Elaborar avaliação do conhecimento sobre o assunto ministrado;
-Uso de computadores para implementar exemplos;

AVALIAÇÃO*

Instrumento(s)	Data
<ul style="list-style-type: none"> Avaliação única por meio do instrumento <i>Quiz</i>; Criação de um projeto final “<i>End to End</i>” implementado em código-fonte reproduzível <i>python</i>. 	Ao fim do módulo estudado.

* No Exame Quiz não é permitido qualquer espécie de acréscimo de nota (trabalhos, participações e outros) estranhos à avaliação.

* Quiz é um jogo mental no qual os jogadores tentam responder corretamente a questões que lhes são colocadas. A palavra também é utilizada como sinônimo de avaliação de aquisição de conhecimentos ou capacidades em ambientes de aprendizagem.

BIBLIOGRAFIA

BIBLIOGRAFIA BÁSICA (Títulos, periódicos, etc)

Título/Periódico	Autor	Edição	Editores	Ano	Livro Texto
Álgebra Linear	José Luiz Boldrini [et al.]	3ª edição	Harbra Ltda	1980	-
Trilhas Python	Eduardo Pereira	1ª edição	Casa do Código	2018	-
Python para Desenvolvedores	Luiz Eduardo Borges	3ª edição	Novatec	2014	Sim

BIBLIOGRAFIA COMPLEMENTAR (Títulos, periódicos, etc)

Título/Periódico	Autor	Edição	Editores	Ano
Curso Intensivo de Python	Eric Matthes	1ª edição	Novatec	2017
Data Science from Scratch	Joel Grus	1ª edição	O'Reilly	2015
Introduction to Applied Linear Algebra	Stephen Boyd	1ª edição	Cambridge University Press	2018

Outros: