Ministério da Educação
UNIVERSIDADE FEDERAL DO PARANÁ
Setor de Ciências Exatas
Departamento de Informática

Ficha 2 (variável)

Disciplina: Ciência de Dados para Segurança						Código: CI1030	
Natureza: ( ) Obrigatória (x) Optativa	(x) Semestral ( ) Anual ( ) Modular						
Pré-requisito: **CI1171**	Co-requisito:	Modalidade: (x) Presencial ( ) Totalmente EaD ( )____*C.H.EaD					
CH Total: 60 **CH semanal: 4**	Padrão (PD): 30	Laboratório (LB): 30	Campo (CP): 00	Estágio (ES): 00	Orientada (OR): 00	Prática Específica (PE): 00	
Estágio de Formação Pedagógica (EFP):	Extensão (EXT): 00	Prática como Componente Curricular (PCC): 00					

Indicar a carga horária semestral (em PD-LB-CP-ES-OR-PE-EFP-EXT-PCC)
***Indicar a carga horária que será à distância.**

EMENTA (Unidade Didática)

Python (Introdução)
Processo de Ciência de Dados
Seleção de técnicas
Estudos de caso e aplicações

PROGRAMA (itens de cada unidade didática)

1. Python: introdução à linguagem de programação Python 3, principais bibliotecas da linguagem para uso em “data science” (numpy, pandas, matplotlib)

2. Processo de ciência de dados: Repositórios de datasets, coleta de dados, extração de características, pré-processamento e limpeza, exploração dos dados e descoberta de conhecimento, modelagem, aplicação de técnicas de mineração de dados e aprendizado de máquina, visualização, tomada de decisões, métricas

3. Seleção de técnicas: seleção de atributos, revisão de algoritmos clássicos de machine learning, escolha adequada de algoritmos para classificação e agrupamento, validação treino/teste, avaliação de resultados, métricas, rotulação

4. Estudos de caso e aplicações: discussão de conceitos-chave sobre ciência de dados para segurança, apresentação de exemplos práticos com dados reais fazendo uso ferramentas livres e de código aberto, principais erros em data science, adversarial machine learning

OBJETIVO GERAL

Apresentar de maneira prática o processo completo de ciência de dados, com foco em descoberta de conhecimento aplicada a dados de segurança.

OBJETIVO ESPECÍFICO

1. Aprender/revisar a linguagem Python
2. Compreender a utilização do processo de ciência/mineração de dados e seus impactos em segurança
3. Coletar, criar, processar e utilizar datasets
4. Revisar técnicas clássicas de aprendizado de máquina e aplicações para segurança
5. Entender, a partir de exemplos, os principais erros cometidos durante o processo de ciência de dados, principalmente em dados específicos no contexto de segurança computacional
6. Modelar um classificador e aplicar na resolução de um problema de segurança

PROCEDIMENTOS DIDÁTICOS

A disciplina será desenvolvida mediante aulas expositivas para apresentação dos conteúdos curriculares teóricos ou demonstrações feitas pelo professor, bem como atividades práticas em laboratório.

As atividades e testes complementares serão disponibilizados via Moodle.

FORMAS DE AVALIAÇÃO

1 prova com peso = 40%
1 projeto final a ser apresentado pelo estudante ou grupo (peso = 60%)

Média Final: 0,3*Prova + 0,7*Projeto

BIBLIOGRAFIA BÁSICA (mínimo 03 títulos)

https://github.com/PenseAllen/PensePython2e

https://jakevdp.github.io/PythonDataScienceHandbook/

Ceschin, F., Oliveira, L. E. S., Grégio, A. R. A. Aprendizado de Máquina para Segurança: Algoritmos e Aplicações. Capítulo 2 do Livro de Minicursos do XIX Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2019), p. 41-90. Disponível em: https://sbseg2019.ime.usp.br/minicursos.pdf

BIBLIOGRAFIA COMPLEMENTAR (mínimo 05 títulos)

https://www.deeplearning.ai/machine-learning-yearning/

http://greenteapress.com/thinkstats2/html/index.html

https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers

https://christophm.github.io/interpretable-ml-book/

http://openbookproject.net/thinkcs/python/english3e/

Professor da Disciplina: André Ricardo Abed Grégio

Assinatura: _____

Chefe de Departamento ou Unidade equivalente: Fabiano Silva

Assinatura: _____

Ministério da Educação
UNIVERSIDADE FEDERAL DO PARANÁ
Setor de Ciências Exatas
Departamento de Informática

CRONOGRAMA – 2021-2

Conteúdo	Horas
Apresentação do Curso	2
Introdução ao Python 3	4
Laboratório	2
Processo de ciência de dados	2
Ciência de Dados x Cibersegurança	2
Numpy/Pandas para processamento de dados	4
Repositórios e datasets	2
Tipos de Dados de Segurança	4
Distribuição e discussão do Projeto	2
Exploração de dados	2
Pré-processamento de dados	2
Modelagem/Representação de dados	4
Prova 1	2
Aprendizado de máquina e segurança	4
Erros comuns/escolha de algoritmos	2
Seleção de atributos/estudos de caso	2
Validação e Testes	2
Visualização/apresentação de dados	2
Tira-dúvidas sobre o Projeto	2
Ajuste fino de classificadores	2
Finalização do Projeto em Laboratório	4
Apresentação dos Projetos	6

Exame Final