

## PLANO DE ENSINO

### Ciência de Dados

#### I – Ementa

Apresentar ao aluno uma visão geral sobre os principais conceitos, técnicas e ferramentas referentes à ciência de dados e dos diversos paradigmas e algoritmos existentes de aprendizado de máquina e inferência indutiva para a aquisição automática de conhecimento a partir de conjuntos de dados.

#### II – Objetivos gerais

Fornecer os principais conceitos, técnicas e ferramentas referentes à ciência de dados e big data, e assim prover teoria básica para que os alunos possam aplicar as novas técnicas e ferramentas estudadas em problemas reais frente à grande quantidade de dados gerados por diferentes fontes.

#### III – Objetivos específicos

Entender os processos de descoberta de conhecimento em bases de dados, mineração e preparação dos dados, pré-processamento, modelagem dos dados, planejamento e análise dos resultados.

#### IV – Competências

Compreender os principais conceitos, técnicas e ferramentas referentes à ciência de dados, para utilização nos processos de mineração e preparação dos dados, pré-processamento, modelagem dos dados, planejamento e análise dos resultados.

#### V – Conteúdo Programático

##### **Introdução e visão geral sobre ciência de dados**

- Big data e ciência de dados – além do hype
- Atual cenário data-driven
- Problemas e desafios
- Soluções baseadas em dados
- Conjuntos de habilidades do profissional cientista de dados
- Perspectivas de um projeto de ciência de dados

##### **Descoberta de conhecimento em bases de dados**

- Visão geral sobre KDD (do inglês, *Knowledge Discovery in Databases*)
- Representação e extração de conhecimento
- Fontes de dados

##### **Introdução e visão geral sobre aprendizado de máquina**

- O que é *Machine Learning* (ML)?
- Modelagem
- *Pipeline* da aprendizagem do modelo
- *Overfitting* e *underfitting*
- Balanço entre viés e variância em modelos de ML

##### **Viés indutivo**

- Sistema de aprendizado
- Tipos de aprendizagem
- Espaço de hipóteses

- Viés de busca – ajuste aos dados

### **Aprendizado descritivo e preditivo**

- Aprendizado supervisionado
- Classificação
- Regressão
- Aprendizado não supervisionado
- Agrupamento
- Associação
- Sumarização

### **Mineração de dados**

- Visão geral
- Modelos de ML
- Árvore de decisão
- Naive-Bayes
- K-Vizinhos mais próximos (KNN)
- K-Médias

### **Preparação de dados**

- Principais fontes de dados
- Coleta, limpeza e organização das informações
- Métodos de raspagem
- Tabulação

### **Pré-processamento de dados**

- Seleção de atributos
- Engenharia de características
- Normalização dos dados
- Dados ausentes

### **Modelos preditivos**

- Regressão linear simples
- Ajuste com mínimos quadrados
- Gradiente descendente
- Regressão linear múltipla
- Regressão logística

### **Planejamento de experimentos**

- *Split* de dados – treino, teste e validação
- Validação cruzada
- *Benchmarking*

### **Análise de resultados experimentais**

- Métricas
- Classificação
- Regressão
- Seleção de modelos

### **Aplicações avançadas de ML**

- Visão computacional
- Processamento de linguagem natural
- Reconhecimento de fala
- APIs de inteligência artificial

## **VI – Estratégias de trabalho**

A disciplina é ministrada por meio de aulas expositivas, metodologias ativas e diversificadas apoiadas no plano de ensino. O desenvolvimento dos conceitos e conteúdos ocorre com o apoio de propostas de leituras de livros e artigos científicos básicos e complementares, exercícios, discussões em fórum e/ou *chats*, sugestões de filmes, vídeos e demais recursos audiovisuais. Com o objetivo de aprofundar e enriquecer o domínio dos conhecimentos e incentivar a pesquisa, o docente pode propor trabalhos individuais ou em grupo, palestras, atividades complementares e práticas em diferentes cenários, que permitam aos alunos assimilarem os conhecimentos essenciais para a sua formação.

## **VII – Avaliação**

A avaliação é um processo desenvolvido durante o período letivo e leva em conta todo o percurso acadêmico do aluno, como segue:

- acompanhamento de frequência;
- acompanhamento de nota;
- desenvolvimento de exercícios e atividades;
- trabalhos individuais ou em grupo;
- Projeto Integrado Multidisciplinar;
- estudos disciplinares;
- atividades complementares.

A avaliação presencial completa esse processo. Ela é feita no polo de apoio presencial no qual o aluno está matriculado, seguindo o calendário acadêmico. Estimula-se a autoavaliação por meio da autocorreção de exercícios, questionários e atividades, de modo que o aluno possa acompanhar sua evolução e rendimento escolar, possibilitando, ainda, a oportunidade de melhoria contínua por meio de revisão e *feedback*.

## **VIII – Bibliografia**

### **Básica**

FACELI, K.; LORENA, A. C.; GAMA, J.; DE CARVALHO, A. C. P. L. F. *Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina*. 1. ed. Rio de Janeiro: LTC, 2011.

GRUS, J. *Data Science do Zero: Primeiras Regras com Python*. 1. ed. Rio de Janeiro: Alta Books, 2016.

MITCHELL, T. M. *Machine Learning*. São Paulo: McGraw-Hill, 1997.

### **Complementar**

ALPAYDIN, E. *Introduction to Machine Learning*. MIT Press, 2004.

FLACH, P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press, 2012.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3. ed.

MORGAN, K. *Series in Data Management Systems*. Morgan Kaufmann, 2011.

PROVOST, F.; FAWCETT, T. *Data Science for Business: What you need to know about data mining and data-analytic thinking* by O'Reilly Media, 2013.

TAN, P. N.; STEINBACH, M.; KUMAR, T. *Introduction to Data Mining*. Addison Wesley, 2005.

WITTEN, I.; FRANK, E. *The Morgan Kaufmann Series in Data Management Systems*. 3. ed. 2011.