



## Plano de Ensino

### 1. Dados de Identificação

Curso: Ciência da Computação Turno: Matutino  
Componente Curricular: Tópicos Especiais em Banco de Dados (optativo)  
Fase: Nona  
Ano/Semestre: 2015/1  
Numero de Créditos: 4  
Carga horária - Hora Aula: 72  
Carga horária - Hora Relógio: 60  
Professor: Denio Duarte  
Atendimento ao aluno: quintas-feiras das 9:00h às 10:00h e sextas-feiras das 10:15h às 11:15h.

### 2. Objetivo Geral do Curso

O curso tem por objetivo a formação integral de novos cientistas e profissionais da computação, os quais deverão possuir conhecimentos técnicos e científicos e serem capazes de aplicar estes conhecimentos, de forma inovadora e transformadora, nas diferentes áreas de conhecimento da Computação. Adicionalmente, os egressos do curso deverão ser capazes de adaptar-se às constantes mudanças tecnológicas e sociais, e ter uma formação ao mesmo tempo cidadã, interdisciplinar e profissional

### 3. Ementa

Mineração de Dados. Aprendizado de Máquina. Modelos de Banco de Dados NoSQL. Map Reduce. Aplicação de mineração de dados em SGBD NoSQL.

### 4. Objetivo

#### 4.1 Geral

- Entender o processo de mineração de dados e seus algoritmos, principalmente aqueles da área de aprendizado de máquina. Aplicar tais conceitos em grandes volumes de dados gerenciados por SGBDs NoSQL.

#### 4.2 Específicos

- Apresentar aos estudantes técnicas de mineração de dados utilizando aprendizado de máquina.
- Capacitar os estudantes a utilizarem algoritmos de aprendizado de máquina para minerar dados.
- Capacitar os estudante na utilização de um SGBD NoSQL
- Capacitar os estudantes no uso da técnica Map Reduce para minerar banco de dados NoSQL

### 5. Cronograma e Conteúdo Programático

Semana	Horas	Total Parc.	Assunto
24,26 02	5	5	Introdução à disciplina Mineração de dados (introdução) Aprendizado de máquina (introdução)
03,05 03	5	10	Avaliação 1 ( $\Lambda_1$ ) Regressão Linear com uma variável Gradiente descendente
10, 12 03	5	15	Avaliação 2 ( $\Lambda_1$ ) Octave
17, 19 03	5	20	Regressão Linear com várias variáveis



Semana	Horas	Total Parc.	Assunto
24,26 03	5	25	Avaliação 3 ( $\Lambda_1$ ) Regressão Logística
31 03	2	27	Equação Normal <b>Feriado</b>
07,09 04	5	32	Regressão Polinomial Avaliação 4 ( $\Lambda_1$ )
14 04	2	34	Regularização <b>ERBD</b>
23 04	3	37	<b>Feriado</b> Regularização
28 04	2	39	Over e Under Fit <b>Feriado</b>
05, 07 05	5	44	Multiclassificador Avaliação 1 ( $\Lambda_2$ )
12, 14 05	5	49	Verificação de hipóteses (precisão e revocação) Algoritmos não supervisionados
19, 21 05	5	54	Clusterização
26, 28 05	5	59	Avaliação 2 ( $\Lambda_2$ ) SGBD NoSQL
02/06	2	61	MongoDB <b>Feriado</b>
09,11 06	5	66	Map Reduce Avaliação 3 ( $\Lambda_2$ )
16, 18 06	5	71	Itens frequentes e regras de associação Avaliação 4 ( $\Lambda_2$ )
23/06	1	72	Apresentação trabalho final

\* O plano e cronograma podem ser alterados pelo professor ao longo do semestre.

## 6. Procedimentos Metodológicos (estratégias de ensino, equipamentos, entre outros)

Conduzir a disciplina com aulas expositivas/dialogadas enquanto discutidos os itens de cunho teórico, evoluindo em tópicos específicos para exercícios práticos, demonstrações, contextualização baseada em publicações atualizadas. Uso de atividades em laboratórios com o objetivo de apresentar/exercitar os conceitos estudados.

O plágio e a cola serão tratados de forma rígida (nota 0 para os envolvidos na atividade em questão).

Não será permitido o uso de equipamentos eletrônicos (computadores, smart-phones, tablets, etc) que não sejam explicitamente para uso na disciplina. O não atendimento a esta regra será passível de punição para toda a turma.

## 7. Avaliação do Processo Ensino-Aprendizagem

Uso de abordagens tais como: provas teóricas, avaliação escrita em aula, exercícios extra-classe, trabalhos de implementação, entre outros.

As avaliações ( $\Lambda$ ) ocorrerão durante o semestre e serão divididas em dois grupos ( $\Lambda_1$  e  $\Lambda_2$ ) e serão calculadas as médias aritméticas de cada grupo ( $\bar{\Lambda}_1$  e  $\bar{\Lambda}_2$ ), sendo que  $\bar{\Lambda}_1$  terá peso 4 (corresponde às avaliações até a metade do semestre) e  $\bar{\Lambda}_2$  terá peso 6 (corresponde às avaliações da metade seguinte) na média final das avaliações (*MFA*), ou seja:

$$MFA = \bar{\Lambda}_1 * 0,4 + \bar{\Lambda}_2 * 0,6.$$

Os trabalhos (*T*) serão também somados e gerada a média aritmética final (*MFT*):

$$MFT = \left( \frac{1}{n} \sum_{i=1}^n T_i \right) \times \kappa$$

onde *n* é o número de trabalhos realizados e  $\kappa$  é o fator de ajuste.  $\kappa=1$  e a cada trabalho não entregue sofre um ajuste



$\kappa = \kappa - 0,05$ .

O cálculo da média final será:

$$MF = MFA * 0,5 + MFT * 0,5.$$

## 7.1 Recuperação: novas oportunidades de aprendizagem e avaliação

As avaliações e trabalhos serão discutidos em sala de aula após a correção dos mesmos. Esta discussão tem como objetivo oferecer uma nova oportunidade de aprendizagem do conteúdo avaliado. Todos os estudantes participarão da discussão.

Se algum estudante obter uma média parcial (MP) abaixo 6,0, será oferecida uma oportunidade, através de uma prova PR com todo o conteúdo, que será utilizada para calcular a nova média como:

$$MP' = MP * 0,5 + PR * 0,5$$

## 8. Referências

### 8.1 Básicas

MEIRA Jr, W., ZAKI, M. J. Data Mining and Analysis: Fundamental Concepts and Algorithms. New York. Cambridge University Press, 2014.

SADALAGE, P. J., FOWLER, M. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. New York. Addison-Wesley, 2013.

MITCHELL, T. Machine Learning. New York, McGraw-Hill Press, 1997

McCREARY, D., KELLY, A. Making Sense of NoSQL: A guide for managers and the rest of us. Manning Publications Co. Shelter Island, NY, 2014.

### 8.2 Complementares

DE AMO S.: Curso Introdutório de Mineração de Dados - Compilação de Notas de Aulas, 2006, 201 páginas. Disponível em [www.deamo.prof.ufu.br](http://www.deamo.prof.ufu.br)

MICHALSKI, R. S., BRATKO, I., KUBAT, M. Machine Learning and Data Mining. Chichester, John Wiley & Sons Ltd, 1998.

TAN, P.-N., STEINBACH, M., KUMAR, V. Introduction to Data Mining. New York, Addison-Wesley, 2006.

STRAUCH, C. NoSQL Databases. Lecture Notes. Stuttgart Media University, 2011. Disponível em [www.christof-strauch.de/nosql dbs.pdf](http://www.christof-strauch.de/nosql dbs.pdf)