



# Big Data

## Curso de Extensão

**Prof. Antonio Guardado**

**Prof. Carlos Eduardo Menezes**

## **Motivação**

- Crescimento exponencial em volume e variedade de dados não pode ser processado e analisado com as ferramentas e técnicas tradicionais
- Necessidade urgente de informação sobre os negócios para tomada de decisão rápida -> vantagem competitiva
- Utilização de técnicas matemáticas e estatísticas em associadas a algoritmos de inteligência artificial para a mineração e análise dos dados
- Carência de profissionais qualificados no mercado de TI : demanda muito maior que a capacidade de formação com expertise e qualidade : tratamento, análise, interpretação e uso das ferramentas adequadas
- Tecnologias emergentes



## **Objetivos**

- Fornecer uma visão geral do assunto
- Não é formar ou preparar especialistas : existem várias áreas da computação, matemática e negócios a serviço do Big Data, demanda uma formação muito extensa e aprofundada em cada área do conhecimento
- Apresentar os principais conceitos, técnicas e ferramentas utilizados
- Compreender os impactos da análise e interpretação dos dados nos negócios

## **Iniciativa**

- **Comissão de Extensão e Pesquisa (CEPE) da Fatec Ipiranga**
- **Presidente : Profa. Dra. Ana Cláudia M. Tiessi G. de Oliveira**
- **Grupo de Pesquisa em Big Data**
  - **Prof. Me. Antonio Guardado ( Líder)**
  - **Prof. Me. Carlos Eduardo Menezes**
  - **Profa. Dra. Andreia Machion**
  - **Profa. Dra. Ana Paula G. Serra**
  - **Prof. Me. Rodrigo Bossini**
  - **Profa. Dra. Ana Cláudia Tiessi**



## **Tópicos**

- **1 – Introdução a Big Data e Business Intelligence (BI)**
- **2 – Banco de dados Relacional (SQL)**
- **3 – Banco de Dados Não-Relacional (NoSQL)**
- **4 – Inteligência Artificial com foco no Aprendizado Automático**
- **5 – Data Mining**
- **6 – Tecnologia para Big Data**



# 1 – Introdução a Big Data e Conceitos

- Dado x Informação



## 2- Banco de Dados Relacional

- **Modelo Relacional de dados**
- **DataWarehouse**
- **DataMining**
- **Consultas em grandes volumes de dados**
- **Limitações dos BDs Relacionais**

### 3- Bancos de dados Não-Relacionais

- Características NO-SQL
- Tipos de Bancos de dados NO-SQL
- Comparação com os BDs Relacionais
- NO-SQL MongoDB
- NO-SQL Neo4J
- Consultas
- Data Mining em NO-SQL





## 4- Inteligência Artificial com Aprendizado de Máquina

- **Aprendizado Supervisionado**
  - Classificação, Previsão, Regressão
- **Aprendizado Não- Supervisionado**
  - Mapas auto-organizáveis, mapeamento do vizinho mais próximo, agrupamento k-means e decomposição em valores singulares
- **Aprendizagem profunda**
  - Redes Neurais

## 5 – Data Mining – Mineração de dados

- **Análise Descritiva**
  - **Análise Prévia**
  - **Análise de outliers ou detecção de desvios**
  - **Visualização**
  - **Descobrimento**
  - **Classificação, Análise de Associações, Agrupamento, Detecção de Sequências, etc.**
- **Análise de Prognóstico**
  - **Predição**
  - **Estimação**
  - **Classificação**



## 6 – Tecnologia para Big Data

- **Processamento Distribuído e Paralelo**
- **Computação em Nuvem (Cloud Computing)**
- **Clusterização**
- **Hadoop**
- **MapReduce**
- **Spark**

# ■ ■ ■ Avaliação





## Bibliografia

- Norvig, P.; Russell, S. Inteligência Artificial. Editora Prentice Hall. 1994
- Richert, W.; Coelho, L. Building Machine Learning Systems with Python. Packt Publishing. 2013.
- O'Neil, C.; Schutt, R. Doing Data Science. O'Reilly, 2013.
- Vaish, G. Getting Started with NoSQL. Packt Publishing. 2013.
- Gupta, A. Learning Apache Mahout. Packt Publishing. 2015.
- Karanth, S. Mastering Hadoop. Packt Publishing. 2014.
- Sinha, S. Making Big Data Work for Your Business. Impact Publishing. 2014.
- Idris, I. Python Data Analysis. Packt Publishing. 2014.
- Sadalage, Pramod J.; Fowler, Martin. NOSQL Essencial , Editora Novatec, 2013