



UNIVERSIDADE ESTADUAL DE MARINGÁ
PRÓ-REITORIA DE ENSINO

PROGRAMA DE DISCIPLINA

Curso:	Programa de Pós-Graduação em Bioestatística		
Modalidade:	Mestrado e Doutorado		
Departamento:	PBE		
Centro:	De Ciências Exatas		
COMPONENTE CURRICULAR			
Nome: Mineração de Dados Massivos			Código: DES_____
Carga Horária: 30 horas	Crédito(s): 2	OPTATIVA	Ano Letivo: 2025

1. EMENTA

Histórico da análise em Big Data, termos e definições. Tecnologias para processamento de Big Data. Métodos de estimação, inferência e de aprendizado para dados massivos.

2. OBJETIVOS

Apresentar conceitos, tecnologias e métodos estatísticos aplicados à mineração de dados massivos, permitindo ao estudante compreender os desafios e aplicar técnicas adequadas de análise, estimação e inferência no contexto de Big Data.

3. CONTEÚDO PROGRAMÁTICO

1. Conceitos Iniciais

1.1. Definições de Big Data

1.2. Dados Estruturados e não estruturados

1.3. Desafios Estatísticos

2. Tecnologias em processamento de dados massivos

2.1. Processamento Paralelo

2.2. Processamento Distribuído

2.3. Apache Hadoop

2.4. Apache Spark

2.5. Processamento de dados massivos em Linguagem R

3. Métodos de estimação e inferência para dados massivos

3.1. Métodos de subamostragem

3.2. P-valor e D-Valor

3.3. Métodos de Monte Carlo para dados massivos

4. Análise Estatística em Computação em Nuvem



UNIVERSIDADE ESTADUAL DE MARINGÁ
PRÓ-REITORIA DE ENSINO

PROGRAMA DE DISCIPLINA

Curso:	Programa de Pós-Graduação em Bioestatística		
Modalidade:	Mestrado e Doutorado		
Departamento:	PBE		
Centro:	De Ciências Exatas		
COMPONENTE CURRICULAR			
Nome: Mineração de Dados Massivos			Código: DES_____
Carga Horária: 30 horas	Crédito(s): 2	OPTATIVA	Ano Letivo: 2025

4. METODOLOGIA

A disciplina será ministrada em modalidade presencial e remoto, com aulas expositivas teóricas e práticas. Serão realizados estudos de caso e exercícios em laboratório de informática, utilizando softwares e ambientes de processamento de dados massivos. Atividades complementares serão propostas para consolidar a aprendizagem.

5. AVALIAÇÃO

A avaliação será composta por exercícios práticos, trabalhos aplicados e/ou estudos de caso, visando verificar a compreensão conceitual e a capacidade de aplicação dos métodos e tecnologias para dados massivos.

6. REFERÊNCIAS

6.1- Básicas (Disponibilizadas na Biblioteca ou aquisições recomendadas)

DEAN, J.; GHEMAWAT, S. *MapReduce: Simplified Data Processing on Large Clusters*. Communications of the ACM, v. 51, n. 1, p. 107–113, 2008.

EFRON, B.; HASTIE, T. **Computer Age Statistical Inference**. Cambridge University Press, 2016.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R**. Springer, 2013.

WHITE, T. **Hadoop: The Definitive Guide**. 4. ed. O'Reilly Media, 2015.

ZAHARIA, M. et al. *Apache Spark: A Unified Engine for Big Data Processing*. Communications of the ACM, v. 59, n. 11, p. 56–65, 2016.

*Disciplina aprovada por “ad referendum”.