



DADOS DE IDENTIFICAÇÃO E ATRIBUTOS¹

CÓDIGO							NOME						DEPARTAMENTO OU EQUIVALENTE					
MATF58							Introdução a bug data						Departamento de Estatística					
CARGA HORÁRIA (estudante)							MODALIDADE/ SUBMODALIDADE						PRÉ-REQUISITO (POR CURSO)					
T	T/P	P	PP	Ext	E	TOTAL	Disciplina/Teórico-prática											
17	0	17	0	0	0	34												
CARGA HORÁRIA (docente/turma)							MÓDULO ²						SEMESTRE DE INÍCIO DA VIGÊNCIA					
T	T/P	P	PP	Ext	E	TOTAL	T	T/P	P	PP	Ext	E	08/09/2020 a 18/12/2020					
17	0	17	0	0	0	34	30	0	30	0	0	0						

EMENTA

Esta disciplina abordará temas de pesquisa atuais relacionados à Estatística e Ciência de Dados, Aprendizado de Máquina e Big Data. Trata-se de disciplina com ementa variável, sendo direcionada à temas recentes que não estejam inicialmente previstos no elenco de disciplinas.

OBJETIVOS

OBJETIVO GERAL

Conhecer de forma prática e teórica os tópicos do programa para uso nas situações relacionadas com a Ciência de Dados e em disciplinas afins.

OBJETIVOS ESPECÍFICOS

Aplicar métodos de análise estatística no contexto de Big Data, bem como conhecer e utilizar as ferramentas necessárias.

CONTEÚDO PROGRAMÁTICO

- Introdução a Análise Preditiva e Analytics
 - 1.1. Business Intelligence, Business Analytics e Big Data
 - 1.2. Dados Estruturados e não estruturados
 - 1.3. Desafios Estatísticos
- Tecnologias em processamento em paralelo
 - 2.1. Sistema de arquivos distribuído e MapReduce
 - 2.2. Apache Hadoop
 - 2.3. Apache Spark
- Métodos de estimação e inferência em processamento em paralelo
 - 3.1. Métodos de Monte Carlo para Big Data
 - 3.1 Métodos de subamostragem
 - 3.2 Seleção de variáveis e regularização
- Análise Estatística em Computação em Nuvem

¹ Os “dados de identificação e atributos” devem estar registrados conforme especificado no Programa do Componente Curricular e disponível no site da Superintendência Acadêmica (SUPAC). O único campo a ser preenchido nesse tópico do formulário é o que diz respeito ao módulo de vagas ofertadas.

² Conforme Resolução CONSUNI 01/2020 e CAE 01/2020, é possível flexibilizar o disposto na Resolução CONSEPE 02/2009.

METODOLOGIA DE ENSINO-APRENDIZAGEM

1. Web conferências e seminários síncronos.
2. Material complementar a ser lido pelo aluno.
3. Vídeo aulas a serem disponibilizadas sobre os tópicos.
4. Fórum de discussão.
5. Pesquisa e trabalhos em grupos.
6. Resolução de exercícios.

Requisito Mínimo Desejável: Recomenda-se que os participantes tenham cursado a Disciplina de Regressão no Bacharelado em Estatística ou possuam conhecimentos básicos de inferência estatística, regressão e linguagens de programação R e/ou Python.

AVALIAÇÃO DA APRENDIZAGEM

Avaliação contínua mediante a evolução dos tópicos do conteúdo programático. Em suma, uma avaliação ao fim de cada módulo (o aluno deverá desenvolver a cada 3 ou 4 semanas). Todas as avaliações serão compostas por duas partes (síncrona e assíncrona). A parte síncrona se baseará em questões conceituais (50%) e a parte assíncrona a trabalhos práticos e/ou de pesquisa (50%). A nota final da disciplina, será composta pela média aritmética das quatro avaliações.

REFERÊNCIAS

REFERÊNCIAS BÁSICAS

Impressos

BENGFORT, B.; KIM, J. *Análítica de dados com Hadoop: Uma Introdução Para Cientistas de Dados*. São Paulo: Novatec, 2016.

IVANOV, Ivan; SINDEREN, Marten J. van; SHISHKOV, Boris (Ed.). *Cloud computing and services science*. New York: Springer, 2012. xvi, 392 p.

MACHADO, Felipe Nery Rodrigues. *Big Data O Futuro dos Dados e Aplicações*. Editora Saraiva, 2018.

Ebooks

BHATNAGAR, Vasudha. *Big Data Analytics: Second International Conference, BDA 2013, Mysore, India, December 16-18, 2013, Proceedings*. XII, 197 p. 50 illus (Lecture Notes in Computer Science, 0302-9743 ; 8302). ISBN 9783319036892.

DI CIACCIO, Agostino; COLL, Mauro; IBANEZ, Jose Miguel Angulo (Ed.). *Advanced statistical methods for the analysis of large data-sets*. Berlin: Springer, 2012. xiv, 486 p. (Studies in theoretical and applied statistics). ISBN 9783642210372.

REFERÊNCIAS COMPLEMENTARES

Impressos

GOMES, Rodrigo Dias Pinho. *Big Data: desafios à tutela da pessoa humana na sociedade da informação*. Rio de Janeiro: Lumen Juris, 2017.

LURASCHI, Javier; KUO, Kevin; RUIZ, Edgar. *Mastering Spark with R: The Complete Guide to Large-Scale Analysis and Modeling*. O'Reilly Media; 1 edition, 2019.

WORLD INNOVATION SUMMIT FOR HEALTH, 13. 2003, Doha.; HEIBECK, Tracy. *Big data and health: Revolutionizing medicine and public health*. Doha, QAT: Qatar Foundation, 2013. 35 p.

Ebooks

BABU, T. Ravindra; MURTY, M. Narasimha; SUBRAHMANYA, S. V. *Compression schemes for mining Large datasets : a machine learning perspective*. New York: Springer, c2013. xvi, 197 p. (Advances in computer vision and pattern recognition, 2191-6586). ISBN 9781447156079

KOŁODZIEJ, J., Di MARTINO, B., TALIA, D., & XIONG, K. (Eds.). *Algorithms and Architectures for Parallel Processing: 13th International Conference, ICA3PP 2013, Vietri sul Mare, Italy, December 18-20, 2013, Proceedings, Part I*. XXII, 481 p. 213 illus (Lecture Notes in Computer Science, 0302-9743 ; 8285). ISBN 9783319038599.

Teses, dissertações, artigos e relatórios técnicos que apresentem metodologias relativas a Big Data.

Docente(s) Responsável(is) à época da aprovação do Plano de ensino-aprendizagem:

Nome: _____ Assinatura: _____

Nome: _____ Assinatura: _____

Aprovado em reunião de Departamento (ou equivalente): _____ em ____/____/____

Assinatura do Chefe

ANEXO

CRONOGRAMA³

Código e nome do componente:	MATF58 – Introdução a Big Data
Nome do/s docente/s:	Anderson Luiz Ara Souza
Período:	08/09/2020 a 18/12/2020

Data ou período de realização	Unidade Temática ou Conteúdo	Técnicas ou estratégias⁴ de ensino previstas	Atividade/ Recurso⁵	CH Docente⁶	CH Discente⁷
Semana 1 e Semana 2	Apresentação da disciplina. 1.Introdução a Análise Preditiva e Analytics 1.1. Business Intelligence, Business Analytics e Big Data 1.2. Dados Estruturados e não estruturados 1.3. Desafios Estatísticos	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos, Artigos e Vídeos	2:30h	2:30h
Semana 2 e Semana 3	2.Tecnologias em processamento em paralelo 2.1. Sistema de arquivos distribuído e MapReduce 2.2. Apache	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online)	AVA e Google for Education, Fórum, Textos, Artigos e	2:30h	2:30h

³ Esta é uma sugestão de cronograma. A sua adoção é facultativa, sendo possível, a critério do(s) professor(es), adotar outra forma de expressar aspectos temporais e de uso de dispositivos tecnológicos. Para o SLS, recomenda-se pensar a organização do componente em unidades ou temáticas amplas, considerando períodos equivalentes à carga horária de uma ou mais semanas

4 Possibilidades de técnicas e estratégias de ensino-aprendizagem:

Síncronas: Aula dialogada (ao vivo) pelos professores em interatividade com os estudantes; Apresentação de artigos ou temas pelos estudantes com mediação dos professores); Aula invertida (*chat* a partir de texto ou vídeo com mediação dos professores); Chats com pequenos grupos.

Assíncronas: Aula expositiva (preleções feitas pelos professores e gravadas como videoaulas); Aula invertida (fórum de discussão a partir de texto ou vídeo) com mediação dos professores; Discussão de tema (problematizado) com X postagem dos estudantes e mediação dos professores; Cocriação de textos colaborativos pelos estudantes com mediação dos professores; Desenvolvimento de atividades/tarefas pelos estudantes: resenha, confecção de vídeos, modelos, questionários, peças jurídicas, roteiros, guias de estudo, produções artísticas com mediação dos professores.

⁵ As palavras **Atividade** e **Recursos** aqui acompanham a classificação do Moodle. As atividades podem ser: Fórum, chat, wiki, tarefas, jogos, escolha, glossários, base de dados, pesquisa, questionário etc. Os recursos podem ser: arquivo, URL, livro, pasta, rótulo etc.

⁶ Indicar carga horária também de elaboração e realização.

⁷ Indicar o tempo previsto para que o estudante realize a atividade/tarefa.

	Hadoop 2.3. Apache Spark	Atividade assíncrona ao decorrer da semana	Vídeos		
Semana 4	Avaliação 1	Avaliação 1	Prova	1:30h	1:30h
Semana 5 a Semana 7	2.3. R e Spark	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos, Artigos, Vídeos e Plataformas hands-on	5:00h	5:00h
Semanas 7 e 8	3. Métodos de estimação e inferência em processamento em paralelo.	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	2:30h	2:30h
Semanas 8 e 9	3.1. Métodos de Monte Carlo para Big Data	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	2:30h	2:30h
Semanas 9 e 10	3.2 Métodos de subamostragem	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	2:30h	2:30h
Semanas 10 a 11	3.3 Seleção de variáveis e regularização	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	3:00h	3:00h
Semana 12	Avaliação 2	Avaliação 2	Prova	2:00h	2:00h
Semana 12 a 14	4. Análise Estatística em Computação em Nuvem	SEX das 13:00h as 14:00h Atividade síncrona (Explicações e	AVA e Google for Education,	5:00h	5:00h

		dúvidas online) Atividade assíncrona ao decorrer da semana	Fórum, Textos, Artigos, Vídeos e Plataformas hands-on		
Semana 14	Avaliação 3	Avaliação 3	Prova	3:00h	3:00h
Semana 15	Avaliação Final	Avaliação	Atividade de Pesquisa e seminário	2:00h	2:00h