



PUC Minas

Pós-graduação Ciência de Dados e Big Data

Projeto Integrado

Sérgio M. Dias

1º semestre de 2018

Apresentação

- Disciplina
 - APL- Projeto Integrado – *Construção Aplicação Big Data e Analytics*
- Datas:
 - **Primeiro semestre 2018**
 - 28 de Abril
 - 07 de Julho
 - **Segundo semestre 2018**
 - A definir

Ementa

- APL- Projeto Integrado – Construção de Aplicações Big Data e Analytics
 - Revisão de padrões.
 - Aspectos arquiteturais de uma solução de big data e ciência de dados.
 - Problemas comuns de big data e ciência de dados.
 - Visão geral de Produtos em Analytics (Data discovery, Data Services,).
 - Utilização de serviços de dados em nuvem.
 - Projetos aplicativos utilizando big data e ciência de dados.
 - Criação de Data Products.
 - Trabalho de conclusão de curso (TCC)

Aprovação

- Frequência
 - 60%
- Avaliações
 - 60%
 - Trabalho de conclusão de curso (TCC) - **individual**
 - Entrega 1 – proposta
 - 10%
 - Entrega 2 – acompanhamento
 - 10%
 - Entrega 3 – artigo técnico científico, protótipo e apresentação final
 - Artigo 30%
 - Protótipo 35%
 - Apresentação 15%

Trabalho de conclusão de curso (TCC)

- Trabalho de conclusão de curso (TCC)
 - Obrigatoriamente com o desenvolvimento de um produto (protótipo mínimo viável) em *big data analytics*
 - Trabalho individual
 - Oportunidade para aplicar todos os conceitos adquiridos na pós-graduação
 - Oportunidade para elaborar um relatório técnico científico
 - Um tipo de produto frequentemente demandado
 - Oportunidade para elaborar um protótipo
 - Novo produto!

Trabalho de conclusão de curso (TCC)

- Etapas macro:
 - 1) Definir um problema de pesquisa (ou necessidade) em big data analytics
 - 2) Definir hipóteses minimamente complexas
 - 3) Definir as bases de dados necessárias
 - Preferencialmente com dados públicos
 - 4) Implementar solução para coleta dos dados
 - 5) Aplicar todas as técnicas de engenharia de dados necessárias para preparação dos dados
 - 6) Explorar os dados

Trabalho de conclusão de curso (TCC)

- Etapas macro:

- 7) Extrair *insights* iniciais a respeito dos dados e a relação com as hipóteses
- 8) Aplicar técnicas de aprendizado de máquina para extração de conhecimento não trivial
- 9) Apresentar os resultados em formato de produto em big data analytics
- 10) Escrever um relatório técnico
 - Formato de trabalho de artigo técnico científico

Trabalho de conclusão de curso (TCC)

- 9) Apresentar os resultados em formato de produto em big data analytics

- Entregar um produto mínimo
- Algumas pequenas ideias baseadas em shiny
 - <https://gallery.shinyapps.io/EDsimulation/>
 - <https://datasociety.com/kitamba-the-opportunity-project/>
 - <https://gallery.shinyapps.io/NYT-bar-optimizer/>
- Outros exemplos
 - <https://serenatadeamor.org>
 - <https://br.okfn.org>
 - https://play.google.com/store/apps/details?id=br.gov.sp.alesp&hl=pt_BR
 - <http://hackfest.com.br>
 - <https://meudeputado.mobi>

Trabalho de conclusão de curso (TCC)

- 9) Apresentar os resultados em formato de produto em big data analytics

- Entregar um produto mínimo
- Algumas pequenas ideias baseadas em shiny
 - <https://gallery.shinyapps.io/EDsimulation/>
 - <https://datasociety.com/kitamba-the-opportunity-project/>
 - <https://gallery.shinyapps.io/NYT-bar-optimizer/>
- Outros exemplos
 - <https://serenatadeamor.org>
 - <https://br.okfn.org>
 - https://play.google.com/store/apps/details?id=br.gov.sp.alesp&hl=pt_BR
 - <http://hackfest.com.br>
 - <https://meudeputado.mobi>

Não ficar limitado apenas a apresentar dados.
Extrair conhecimento!

Trabalho de conclusão de curso (TCC)

- 10) Escrever um relatório técnico

- Formato de trabalho de artigo técnico científico
- Um tipo de produto em ciência de dados
- Modelo:
 - <http://www.sbc.org.br/documentos-da-sbc/category/169-templates-para-artigos-e-capitulos-de-livros>
- Mínimo de 12 laudas

Trabalho de conclusão de curso (TCC)

- 10) Escrever um relatório técnico

- Formato básico
 - Resumo/Abstract
 - Introdução
 - Motivação
 - Referencial teórico
 - Trabalhos relacionados
 - Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
 - Experimentos
 - Conclusões e trabalhos futuros
 - Referências bibliográficas

Trabalho de conclusão de curso (TCC)

- 10) Escrever um relatório técnico

- Formato básico
 - Resumo/Abstract
 - Introdução
 - Motivação
 - Referencial teórico
 - Trabalhos relacionados
 - Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
 - Experimentos
 - Conclusões e trabalhos futuros
 - Referências bibliográficas

Apresentar a solução tecnológica, incluindo a arquitetura da solução proposta.

Exemplos de trabalhos

- Silva, P. R., Dias, S. M., Brandao, W. C., Song, M. A., & Zárate, L. E.. **Formal Concept Analysis applied to Professional Social Networks Analysis**. In: 19th International Conference on Enterprise Information Systems (ICEIS), 2017, Porto. Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS), 2017. v. 1. p. 123-12.
- Dias, S. M.; Nogueira, B. M.; Zárate, L. E.; Vieira, N. J.. **Extracão e Representação de Conhecimento de Redes Neurais Artificiais Utilizando o Reticulado Conceitual Iceberg e Extração de Regras de Implicacão no Método FCANN**. In: V WAAMD - Workshop em Algoritmos e Aplicações de Mineração de Dados, 2009, Fortaleza, CE. Anais do V WAAMD, 2009. p. 10-17.
- Pita, M. R. de S.; Torres, G. da G.; Dias, S. M.. **Análise de Redes Sociais para Inferência de Similaridade entre Fornecedores de Governo**. In: ConSerpro, 2015, Florianópolis. ConSerpro - Congresso SERPRO de Tecnologia e Gestão Aplicada a Serviços Públicos, 2015. p. 1-30.

Apresentação

- **Formal Concept Analysis applied to Professional Social Networks Analysis .**

Apresentação

- **Extracão e Representação de Conhecimento de Redes Neurais Artificiais Utilizando o Reticulado Conceitual Iceberg e Extração de Regras de Implicacão no Método FCANN.**

Apresentação

- **Análise de Redes Sociais para Inferência de Similaridade entre Fornecedores de Governo.**

Outros exemplos de trabalhos disponíveis

- Gonçalves, W. de B.; Zárate, L. E.; Assis, E. S.; Dias, S. M.. **Estimativa de Dados Climáticos Utilizando Redes Neurais Artificiais para fins de Zoneamento Bioclimático do Estado de Minas Gerais**. In: XII Encontro Nacional de Tecnologia. In: XII Encontro Nacional de Tecnologia, 2008, Fortaleza-CE. XII Encontro Nacional de Tecnologia, 2008. p. 1-10.
- Zárate, L. E.; Dias, S. M.. **Qualitative behavior rules for the cold rolling process extracted from trained ANN via the FCANN method**. Engineering Applications of Artificial Intelligence, v. 2, p. 718-731, 2009.

Entregas

Entrega 1 – proposta- Data 04 de julho de 18

- Resumo/Abstract
- Introdução
 - Motivação
- Referencial teórico
- Trabalhos relacionados
- Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
- Experimentos
- Conclusões e trabalhos futuros
- Referências bibliográficas

Entregas

Entrega 1 – proposta- Data 04 de julho de 18

- Resumo/Abstract
- Introdução
 - Motivação
- Referencial teórico
- Trabalhos relacionados
- Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
- Experimentos
- Conclusões e trabalhos futuros
- Referências bibliográficas

Quais dados ajudam a resolver o(s) problema(s)?

Entregas

Entrega 1 – proposta- Data 04 de julho de 18

- Entregar a proposta
- sergiomariano@gmail.com
- Assunto: "[pos-pucminas-APL] Proposta"
- Preparar apresentação de 10 minutos da proposta
 - Oportunidade para discutir a proposta e obter orientações iniciais
 - Demonstrar o rascunho inicial da arquitetura de solução

Entregas

Entrega 2 – acompanhamento- primeira aula do segundo semestre

- Resumo/Abstract
- Introdução
 - Motivação
- Referencial teórico
- Trabalhos relacionados
- Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
- Experimentos
- Conclusões e trabalhos futuros
- Referências bibliográficas

Entregas

Entrega 2 – acompanhamento- primeira aula do segundo semestre

- Entregar a evolução do trabalho
- sergiomariano@gmail.com
- Assunto: "[pos-pucminas-APL] Evolução"
- Preparar apresentação de 10 minutos
 - Oportunidade para obter orientações sobre o direcionamento do trabalho.
 - Demonstrar o estado atual do produto/protótipo

Entregas

Entrega 3 – artigo técnico científico e apresentação final – Artigo, protótipo e apresentação - segunda aula do segundo semestre

- Resumo/Abstract
- Introdução
 - Motivação
- Referencial teórico
- Trabalhos relacionados
- Nova proposta (Metodologia)
 - Domínio do problema
 - Coleta
 - Pré-processamento
 - Extração de conhecimento
- Experimentos
- Conclusões e trabalhos futuros
- Referências bibliográficas

Entregas

Entrega 3 – artigo técnico científico e apresentação final - Artigo, protótipo e apresentação - segunda aula do segundo semestre

- Entregar o trabalho final
- Link para o protótipo
- Todo código fonte da solução e demais artefatos
- sergiomariano@gmail.com
- Assunto: "[pos-pucminas-APL] Trabalho final"
- Preparar apresentação final de 20 minutos
 - Apresentar o trabalho final e o protótipo

Recursos

- Recursos na web

- Machine Learning, Data Science, Big Data, Analytics, AI
 - <https://www.kdnuggets.com/>
- The Home of Data Science & Machine Learning
 - <https://www.kaggle.com/>
- https://pt.wikibooks.org/wiki/Processamento_de_Dados_Massivos
- <http://posgraduando.com/>
- <http://phdcomics.com/>



Bibliografia

- Baseado em artigos científicos



Obrigado pela atenção!

Sérgio M. Dias | www.sergiomdias.com
Professor e Pesquisador | PUC Minas
<http://www.icei.pucminas.br/projetos/dsrgroup/>
Doutor em Ciência da Computação | UFMG
www.dcc.ufmg.br
Analista Sênior – Cientista de dados | SERPRO
www.serpro.gov.br

Contato: sergiomariano@gmail.com