



DISCIPLINA: TÓPICOS ESPECIAIS: Análise Visual de Dados

SEMESTRE: Semestre Impar

CÓDIGO DA DISCIPLINA: ESPA04-06

CARGA HORÁRIA: 56 HORAS

CRÉDITOS: 4

PROFESSOR: RENATO NOVAIS

De 2018.1 a Atual

PÁGINA DA DISCIPLINA: <http://ppgesp.ifba.edu.br/ESPA0406>

Ementa

Big Data, Análise de Dados, Visualização de Informação, Análise Visual de Dados

Aulas

| id | Data | Assunto | Quem |
|----|------------|---|---------------|
| 1 | 05/07/2018 | Introdução ao curso Big data | Renato Novais |
| 2 | 12/07/2018 | Mineração de Dados | Manoel Neto |
| 3 | 19/07/2018 | Estatística | Renato Novais |
| 4 | 26/07/2018 | Visualização de Dados Estatística | Renato Novais |
| 5 | 02/08/2018 | Visualização de Dados Introdução a Machine Learning | Renato Novais |
| 6 | 09/08/2018 | Apresentação 1 - Trabalho final | Todos alunos |

| | | | |
|----|------------|---|--|
| 7 | 16/08/2018 | Computação paralela/Hadoop Apresentação Mini-trabalhos Acompanhamento Trabalho Final | Francisco Borges Renato Novais |
| 8 | 23/08/2018 | Dados abertos | Pablo Florentino |
| 9 | 30/08/2018 | Visual analytics Computação paralela/Hadoop | Flávio Dusse Francisco Borges |
| 10 | 06/09/2018 | Análise de Sentimentos Testes Adaptativos Computadorizados | Glauçia Boechat Victor Jatobá |
| 11 | 13/09/2018 | Algoritmos genéticos Visual analytics | Frederico Barbosa Flávio Dusse |
| 12 | 20/09/2018 | Apresentação 2 – Trabalho final Algoritmos genéticos | Todos alunos Frederico Barbosa |
| 13 | 04/10/2018 | Acompanhamento Trabalho Final | Todos alunos |
| 14 | 11/10/2018 | Apresentação 2 - Trabalho final | Todos alunos |

Avaliação

1. Mini-Trabalhos (dupla)
 - a. Apresentação de 5 a 10 minutos
 - b. Enviar até o dia anterior da apresentação para o professor
2. Trabalho Final (dupla)
 - a. Escolher uma técnica relacionado a big data
 - b. Identificar pelo menos 5 artigos recentes e conceituados sobre o tema, mostrando o que há de mais recente na área
 - c. Fazer um trabalho de análise de dados prático relacionado ao tema, com uma massa de dados bastante grande
 - d. Utilizar visualização de dados para visualizar a massa de dados
 - e. Fazer uma análise crítica relacionando os benefícios da técnica utilizada e do uso de visualização.
 - f. Escrever um artigo (2 páginas coluna dupla) sobre o trabalho
3. Apresentação 1 - Trabalho final
 - a. Definição da Técnica escolhida (Teoria)
 - b. Apresentação dos artigos escolhidos
 - c. Apresentar a massa de dados
 - d. Apresentar o que pretende fazer com ela
4. Apresentação 2 - Trabalho final

- a. Evoluir a apresentação anterior
- b. Mostrar o uso da técnica utilizada e da visualização dos dados
- c. Apresentar a análise crítica

Possibilidades de Trabalhos

1. Leitura de artigos
2. Uso de técnicas para Big Data + visualização
3. Aplicação de ferramentas

Mini-Trabalhos

| Id | Descrição | Data Apresentação/Entrega |
|----|--|-------------------------------------|
| 1 | Escolher um problema de big data e apresentar ele considerando os 5Vs | 19/07/2018 |
| 2 | Escolher uma ferramenta de data mining, processar uma base de dados externa à ferramenta (i.e. não deve ser os próprios exemplos da ferramenta) e explicar quais técnicas/tarefas de data mining foram utilizadas Exemplos de possíveis ferramentas: https://www.kdnuggets.com/ https://waikato.github.io/weka-wiki/ - Qual base de dado escolheu? como obteve os dados? como pré-processou? qual algoritmo? o que se conseguiu encontrar? | 26/07/2018 02/08/2018 |
| 3 | Para a mesma base de dados do mini-trabalho 2, você deve escolher algum atributo de interesse (ex. Número de vendas, salário), e aplicar estatística descritiva (Média, Mediana, moda, range, variância, desvio padrão, Quartis, Histograma, outliers). Depois disso, você deve dividir os dados amostrais em dois grupos e aplicar teste de hipótese | 16/08/2018 |
| 4 | MiniTrabalho sobre dados abertos: "Acessar um dos portais de transparência / dados abertos apresentados em sala: – Escolher uma coleção de dados entre as | 13/09/18 |

| | | |
|---|---|------------|
| | <p>dimensões utilizadas pelo ODI Brasil</p> <ul style="list-style-type: none"> – Propor uma aplicação de cunho social que utilize estes dados, descrevendo-a e esboçando sua arquitetura. – Leitura de um dos artigos indicados com resenha " <p>Obs: para este exercício, não vale o artigo sobre "datacracia"</p> | |
| 5 | <ul style="list-style-type: none"> - Instalar o Hadoop Single Node Setup - Fazer código para o Hadoop, que em vez de contar palavras, você deve contar o número total de letras para cada palavra de um texto. <p>Requisitos:</p> <ul style="list-style-type: none"> - Você pode usar qualquer arquivo de entrada, mas o arquivo deve ter pelo menos 1000 palavras. Use este site de geração de texto fictício: http://www.blindtextgenerator.com - Você precisa executar seu código no nó único do Hadoop e obter o resultado. <p>Sugestão:</p> <p>Use o Eclipse para codificação e execução</p> | 20/09/2018 |
| 6 | <p>Mini-Trabalho de Análise de sentimentos. Descrição disponível no material da aula no site.</p> | |
| 7 | | |
| 8 | | |

Ferramentas:

1. Big Data
 - a. Apache Spark: <https://spark.apache.org/>
2. Machine Learning
 - a. Tensor Flow: <https://www.tensorflow.org/>
3. Visualização

Links

1. https://www.of-course.me/Online-Courses?w=data+analytics&gclid=EAlalQobChMIs_WvituK2gIVAgWRCh1WDgEJEAAAYBCAAEqKj7PD_BwE
2. http://news.codecademy.com/what-is-machine-learning/?utm_source=customer.io&utm_medium=email&utm_campaign=fortnightly_2-22-18&utm_content=whatismachinelearning
3. Estadística: OpenIntro Stats https://www.openintro.org/stat/textbook.php?stat_book=os
4. Learning python <https://www.codecademy.com/learn/learn-python>
5. Deep Learning with python
<https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=3>
6. Bigdata
 - a. Cursos
<https://www.of-course.me/Big-Data-Fundamentals/course-238600>
 - b. Sites & papers
<https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>