****

**Orientações para a elaboração do Trabalho de Conclusão de Curso - TCC**

**Pós-Graduação em BIG DATA**

**O trabalho de conclusão de Curso será feito individualmente.**

1. **Entregas**

A nota final de TCC é formada por uma média ponderada das notas das três entregas. Caso o aluno tenha alguma viagem pela empresa, o TCC deverá ser entregue antes da data prevista. **Não será aceito de forma alguma a entrega fora do prazo.**

Todas as entregas devem ser enviadas para os e-mails: [TCCBIGDATALABDATA@FIA.COM.BR](mailto:TCCBIGDATALABDATA@FIA.COM.BR) e [joaopcnogueira@outlook.com](mailto:joaopcnogueira@outlook.com) **até as 23h do dia da entrega com os devidos anexos.**

O **assunto do e-mail** deverá ter o seguinte formato:

<CÓDIGO DA TURMA> - <ENTREGA> - <NOME DO ALUNO>

Exemplo:

POST14 - ENTREGA 1 - CAÍQUE TADEU FILIPINI

**E-mails enviados sem o conteúdo do TCC, com arquivos errados e com arquivos incompletos NÃO serão aceitos.**

1. **Tema**

O TCC deverá ser realizado considerando uma das opções:

* **Projeto de Analytics** considerando os seguintes tópicos:
  + Análise Exploratória de Dados
  + Modelagem com Técnicas de Estatística Aplicada
  + Modelagem com Inteligência Artificial (Machine Learning ou Deep Learning)
  + Comparação de modelos
  + Recomendações
* **Projeto de Big Data** considerando os seguintes tópicos:
  + Coleta, ingestão e processamento de dados
  + Solução de visualização dos dados
  + Detalhamento da arquitetura do Data Lake
  + Disponibilização dos dados para análise
  + Análise Exploratória de Dados
  + Opcional - Modelagem com Inteligência Artificial (Machine Learning ou Deep Learning)

O projeto de Big Data deve ser autorizado pelo prof. Daniel, pelo prof. Anderson ou pela profa. Alessandra

1. **Critério para Aprovação**

A nota final será obtida pela nota ponderada das três entregas e deverá ser maior ou igual a sete (7,0) para a aprovação.

Peso da entrega 1 – 5 % (29/05/2020)

Orientação 1 de TCC: 27/07/2020

Peso da entrega 2 – 30 % (13/08/2020)

Orientação 2 de TCC: 14/09/2020

Peso da entrega 3 – 65 % (22/10/2020)

Caso o aluno não obtenha a nota igual ou superior a 7 será reprovado e terá o direito de fazer um aditivo contratual (por mais 3 meses) mediante o pagamento da dependência.

1. **Conteúdo das Entregas** 
   1. Projeto de Analytics

**Todas as entregas do TCC deverão ser elaboradas em formato de Sumário Executivo realizado em formato de apresentação (ppt) e enviadas em PDF.**

**~~1º Entrega – PESO 5 %~~** ~~(29/05/2020)~~

~~Conteúdo:~~

1. ~~Capa com título do trabalho e data de entrega~~
2. ~~Folha de apresentação contendo: nome do curso, nome do aluno, nome dos coordenadores do curso~~
3. ~~Objetivo do trabalho bem detalhado~~
4. ~~Contextualização do problema e explicação do período de extração dos dados para análise~~
5. ~~Descrição das variáveis da base de dados~~
6. ~~Informar o tamanho da base de dados (quantidade de registros)~~

**~~2º Entrega – PESO 30 % (13/08/2020)~~**

~~Conteúdo:~~

~~1.Capa com título do trabalho e data~~

1. ~~Folha de apresentação contendo: nome do curso, nome do aluno, nome dos coordenadores do curso~~
2. ~~Objetivo do trabalho bem detalhado~~
3. ~~Contextualização do problema~~
4. ~~Análise Exploratória das variáveis~~
   1. ~~Para as variáveis qualitativas fazer tabelas de frequência e para as variáveis quantitativas calcular as medidas de posição (média, quartis), de dispersão (desvio padrão e coeficiente de variação) e Box-plot.~~
   2. ~~Comentar~~ **~~TODAS~~** ~~as tabelas e gráficos.~~
   3. ~~Não colocar a saída do software.~~
   4. ~~As tabelas devem estar no padrão internacional.~~

~~Exemplo:~~

|  |  |
| --- | --- |
| ~~Sexo~~ | ~~Frequência~~ |
| ~~Masculino~~ | ~~0,7~~ |
| ~~Feminino~~ | ~~0,3~~ |

1. ~~Modelagem~~

~~Ajustar todas as~~ **~~técnicas de estatística tradicional~~** ~~que forem compatíveis com o objetivo do trabalho (Regressão linear, regressão logística, cluster e árvore de decisão)~~

~~Comentar os resultados obtidos~~

1. ~~Conclusões~~

~~Obs: copiar os códigos dos programas e as saídas dos softwares em um arquivo e enviar na entrega 3.~~

**3º Entrega – Data: PESO 65 %**

Conteúdo:

1.Capa com título do trabalho e data

1. Folha de apresentação contendo: nome do curso, nome do aluno, nome dos coordenadores do curso
2. Objetivo do trabalho bem detalhado
3. Contextualização do problema
4. Análise Exploratória das **principais variáveis**
   1. Para as variáveis qualitativas fazer tabelas de frequência e para as variáveis quantitativas calcular as medidas de posição (média, quartis), de dispersão (desvio padrão e coeficiente de variação) e Box-plot.
   2. Comentar **TODAS** as tabelas e gráficos das principais variáveis.
   3. Não colocar a saída do software.
   4. As tabelas devem estar no padrão internacional.

Exemplo:

|  |  |
| --- | --- |
| Sexo | Frequência |
| Masculino | 0,7 |
| Feminino | 0,3 |

1. Modelagem

* Ajustar todas as **técnicas de estatística tradicional** que forem compatíveis com o objetivo do trabalho (Regressão linear, regressão logística, cluster e árvore de decisão).
* Ajustar todas as **técnicas de inteligência artificial, apresentadas no curso**, que forem compatíveis com o objetivo do trabalho.
* Apresentar as métricas de qualidade de ajuste dos modelos.
* Comentar os resultados obtidos comparando as metodologias.

1. Conclusões

Apresentar as conclusões do trabalho

1. Enviar um arquivo com todos os códigos dos programas e as saídas dos softwares.
   1. **Projeto de Big Data**

**1º Entrega – Data: PESO 5 %**

**Entrega elaborada em formato de Sumário Executivo, realizada em formato de apresentação (ppt) e enviada em PDF.**

Lembrando que um Projeto de Big Data deve considerar os 7 Vs (Volume, Variedade, Velocidade, Veracidade, Visualização, Vulnerabilidade e Valor) destaque no seu trabalho os 7 Vs.

**Conteúdo:**

1. Capa com título do trabalho e data.
2. Folha de apresentação contendo: nome do curso, nome do aluno, nome dos coordenadores do curso.
3. Objetivo do trabalho bem detalhado. De que forma você vai agregar **VALOR**?
4. Contextualização do problema e explicação do período de extração dos dados para análise.
5. Descrição das variáveis da base de dados. Destaque a **VARIEDADE** dos dados: estruturados, semi-estruturados ou não estruturados.
6. Destacar como você verificará a **VERACIDADE** dos dados.
7. Indicar a **volumetria** estimada.
8. Apresentar o desenho da arquitetura da solução proposta, justificando cada componente do desenho.

Perguntas que você deve responder para ajudar a elaborar o tópico 8:

- Quais ferramentas você vai utilizar para a coleta dos dados? Você vai criar uma ferramenta para ingestão ou vai utilizar alguma(s) que já exista? (Sqoop ou etc..)

- Os dados virão de um banco de dados relacional, NoSql, API, crawler de internet?

- Quais ferramentas você vai utilizar para processamento dos dados? Batch e real time? (Spark ou etc)

- Qual Arquitetura você vai utilizar? (Lambda, Kappa dentre outras)

- Como você vai criar os metadados? (Hive dentre outros)

- Vai utilizar ferramentas intermediárias? (Kafka dentre outras)

- Sistema de arquivos? (S3, HDFS dentre outros)

- Formato de arquivos? (Parquet, OCR dentre outros)

- Algum banco de dados para facilitar a leitura? (HBase, Cassandra ou etc)

**2º Entrega – PESO 30 %**

Nesta entrega é permito alterar o que foi entregue na 1º Entrega. O conteúdo da 1º Entrega pode ser melhorado ou modificado.

Os tópicos de 1 a 8 devem ser elaborados em formato de Sumário Executivo, realizados em formato de apresentação (ppt) e enviados em PDF.

A ingestão de dados (tópico 9) deve estar pronta na data da 2º entrega e será apresentada na data agendada pelo LABDATA.

1.Capa com título do trabalho e data

2.Folha de apresentação contendo: nome do curso, nome do aluno, nome dos coordenadores do curso

3.Objetivo do trabalho bem detalhado. De que forma você vai agregar **VALOR**?

4.Contextualização do problema e explicação do período de extração dos dados para análise

5. Descrição das variáveis da base de dados. Destaque a **VARIEDADE** dos dados: estruturados, semi-estruturados ou não estruturados.

6. Destacar como você verificará a **VERACIDADE** dos dados.

7.Indicar a **volumetria** estimada.

8.Apresentar o desenho da arquitetura da solução proposta, justificando cada componente do desenho

9.Apresentar a ingestão de dados (Apresentar o desenho da solução (item 8) funcionando).

**3º Entrega – PESO 65 %**

Lembrando que um Projeto de Big Data deve considerar os 7 Vs (Volume, Variedade, Velocidade, Veracidade, Visualização, Vulnerabilidade e Valor) destaque no seu trabalho os 7 Vs.

A 3º Entrega deve conter todo o conteúdo da 2º Entrega (É permitido realizar modificações e melhorias no conteúdo da 2º Entrega)

O tópico 1 (abaixo) deve ser elaborado em formato de Sumário Executivo, realizado em formato de apresentação (ppt) e enviado em PDF.

O tópico 2 (abaixo) deve estar pronto na data da 3º entrega e será apresentada na data agendada pelo LABDATA.

1.Apresentar como serão realizadas as tarefas:

* + Coleta, ingestão e processamento de dados
  + Solução de **visualização** dos dados
  + Apresentar o detalhamento da arquitetura do Data Lake
  + Disponibilização dos dados para análise
  + Apresentar a Análise Exploratória de Dados
  + Opcional – Apresentar a Modelagem com Inteligência Artificial (Machine Learning ou Deep Learning)

2. Apresentar:

* + Coleta, ingestão e processamento de dados
  + Solução de visualização dos dados
  + Disponibilização dos dados para análise
  + **Protótipo funcionando - POC**