



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 o email: TCCBIGDATALABDATA@FIA.COM.BR até as 23:00 hs do dia da entrega com os devidos anexos.

O assunto do email deverá ter o seguinte formato: <código da turma>-<Nome do aluno>. Exemplo: MBAT16-DANIEL CASTRO

Emails enviados sem o conteúdo do TCC, com arquivos errados e com arquivos incompletos NÃO serão aceitos.

2. 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

3. 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 %

Peso da entrega 2 – 30 %

Peso da entrega 3 – 65 %

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.

4. Conteúdo das Entregas

4.1 Projeto de Analytics

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

1º Entrega – PESO 5 %

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 %

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
4. Contextualização do problema
5. Análise Exploratória das variáveis
 - a. 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.
 - b. Comentar **TODAS** as tabelas e gráficos.
 - c. Não colocar a saída do software.
 - d. As tabelas devem estar no padrão internacional.

Exemplo:

Sexo	Frequência
Masculino	0,7
Feminino	0,3

6. 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

7. 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
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
5. Análise Exploratória das **principais variáveis**
 - a. 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.
 - b. Comentar **TODAS** as tabelas e gráficos das principais variáveis.
 - c. Não colocar a saída do software.
 - d. As tabelas devem estar no padrão internacional.

Exemplo:

Sexo	Frequência
Masculino	0,7
Feminino	0,3

6. 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.
7. Conclusões

Apresentar as conclusões do trabalho
8. Enviar um arquivo com todos os códigos dos programas e as saídas dos softwares.

4.2 Projeto de Big Data

1º Entrega – Data: PESO 5 %

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
4. Contextualização do problema e explicação do período de extração dos dados para análise
5. Descrição das variáveis da base de dados
6. Indicar a volumetria estimada
7. Cenário para ingestão de dados

2º Entrega – PESO 30 %

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
4. Contextualização do problema e explicação do período de extração dos dados para análise
5. Descrição das variáveis da base de dados
6. Cenário para ingestão de dados
7. Apresentar o desenho da arquitetura da solução proposta, justificando cada componente do desenho
8. Apresentar a ingestão de dados

3º Entrega – PESO 65 %

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

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

5.Descrição das variáveis da base de dados

Apresentar:

- 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)
- Protótipo funcionando - POC