

# Trabalho de Conclusão de Curso

## ORIENTAÇÕES

---

Prezado aluno,

Nesse documento estão apresentadas as informações necessárias para o desenvolvimento do seu Trabalho de Conclusão de Curso. O TCC nos cursos de especialização da PUC Minas Virtual é um trabalho interdisciplinar. Nosso propósito é consolidar os conhecimentos aprendidos no curso, dando ao aluno mais uma oportunidade de colocá-los em prática em um contexto de trabalho.

No TCC você deverá desenvolver um projeto de Ciência de Dados passando por várias etapas, desde a definição do problema até a comunicação dos resultados.

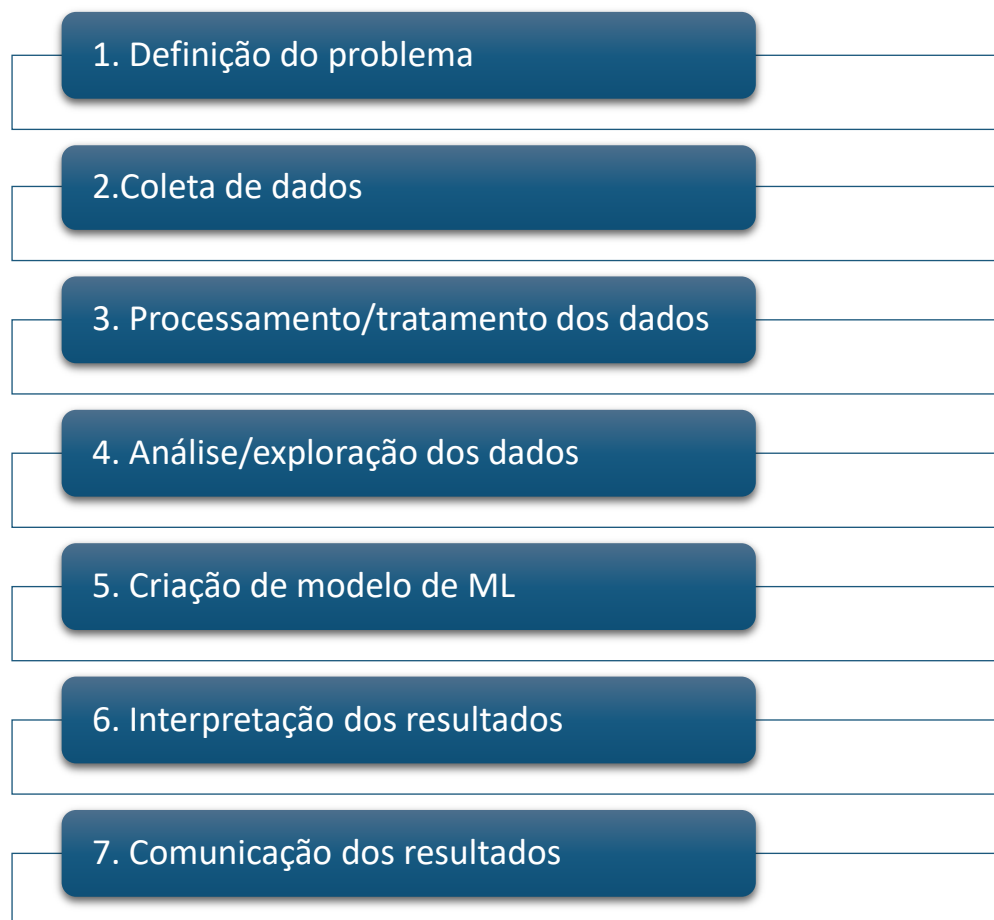
O trabalho pode ser individual ou em grupos de até 3 pessoas.

Se houver qualquer dúvida, envie uma mensagem no Fórum de Discussões da disciplina. Assim, todos os alunos podem ver as perguntas realizadas pelos colegas e os esclarecimentos sobre elas.

## ESCOPO DO TRABALHO

---

O diagrama a seguir motras as etapas do desenvolvimento do seu trabalho.



## **1. Definição do Problema**

No seu TCC, primeiramente você deverá escolher um problema de seu interesse em qualquer área: transporte, economia, consumo, educação, saúde, etc. Você tem liberdade para escolher o tema que achar mais interessante. Pode ser algo relacionado ao seu trabalho ou a bases de dados disponíveis na internet.

## **2. Coleta de Dados**

Como sugestão para obtenção de *datasets*, você pode buscar em sites como o Kaggle, o famoso repositório de dados da Universidade da Califórnia em Irvine (UCI), os dados abertos do governo federal (dados.gov.br) ou dados de outros países (EUA – data.gov, Reino Unido – data.gov.uk). Recentemente, a comunidade R Ladies publicou uma lista bem interessante de sites onde você pode obter dados para projetos de ciência de dados. Para visualizar essa lista, clique [aqui](#).

*Observação: caso você queira trabalhar com dados da empresa onde você trabalha, sugerimos que você remova ou camufle qualquer dado que identifique pessoas ou organizações.*

Para o seu TCC, você deverá utilizar múltiplas fontes de dados. Ou seja, você deve usar mais de um *dataset*, ou utilizar técnicas de Recuperação de Informação para recuperar informações na Web e assim enriquecer sua coleta.

## **3. Processamento/Tratamento de Dados**

Após a obtenção dos seus dados, é momento de tratá-los para que eles possam ser analisados. Aqui você deve utilizar ferramentas que permitam tratar dados ausentes ou duplicados, inconsistências de dados, etc. Lembre-se que você deverá utilizar múltiplas fontes de dados.

## **4. Análise e Exploração dos Dados**

Nessa etapa você começará a explorar seus dados de uma forma mais analítica, tentando elaborar ideias, levantar hipóteses e começando a identificar padrões em seus dados. Talvez você sinta a necessidade de voltar em passos anteriores e obter mais dados e tratá-los para conseguir responder ao problema proposto.

## **5. Criação de Modelos de Machine Learning**

Embora seja muito comum, essa não é uma etapa obrigatória em todos os projetos de Ciência de Dados e, portanto, você não será obrigado a criar/aplicar modelos de Machine Learning para identificar padrões e fazer previsões. Entretanto, encorajamos que você faça alguns modelos, mesmo que mais simples, para fazer agrupamentos de dados, indicar padrões e tendências ou fazer previsões.

## **6. Interpretação dos Resultados**

Nessa etapa você deve interpretar os resultados obtidos na análise e exploração de dados e também interpretar os resultados da aplicação dos modelos de Machine Learning, descobrindo insights importantes para responder o problema proposto.

## **7. Apresentação dos Resultados**

Para encerrar, você deve apresentar o resultados que respondem o problema proposto. Use e abuse de gráficos, tabelas e, principalmente, de sua criatividade, para comunicar seus resultados da forma mais efetiva possível. Um Cientista de Dados também deve ser um bom contador de histórias.

## USO DE UM MODELO CANVAS PARA REGISTRAR SEU WORKFLOW

---

Louis Dorard e Jasmine Vasandani desenvolveram modelos Canvas, baseados no consagrado *Business Model Canvas*, para ajudá-los em seus projetos de Ciência de Dados e *Machine Learning*. Essa é uma boa maneira de programar as etapas do seu projeto e de apresentar um resumo do que foi feito. Você pode obter mais informações sobre o modelo proposto por Louis Dorard em sua página (clique [aqui](#)) e sobre o modelo desenvolvido por Jasmine Vasandani em um artigo publicado no *Towards Data Science* (clique [aqui](#)).

## ENTREGA DO TRABALHO

---

Para realizar a entrega do trabalho, você deve postar o seu TCC conforme o template disponibilizado, em formato PDF, na Atividade Aberta 01 no Canvas, até o dia 08/09/2019. Os links para o vídeo e repositório de dados devem estar contidos nesse documento.

Após avaliação das entregas postadas no Canvas, os professores da banca irão indicar os alunos aptos para a apresentação final do TCC.

## O QUE DEVE SER ENTREGUE?

---

- Relatório conforme template disponibilizado (em formato PDF).
- Link para vídeo de até 5 minutos apresentando de forma sucinta o seu projeto, desde a definição do problema, a obtenção e tratamento dos dados, até a apresentação dos resultados.
- Endereço do repositório (Github, One Drive, Google Drive, Dropbox, etc...) contendo os scripts desenvolvidos e os dados utilizados e gerados em seu trabalho.

## DÚVIDAS?

---

Nosso objetivo foi disponibilizar todos os materiais necessários para a execução do trabalho. Entretanto, entendemos que dúvidas podem surgir.

Neste caso, mande uma mensagem para a gente no fórum de discussão.

Bom trabalho!