

## Orientações sobre o repositório e-gov-data-mining

As equipes dos grupos 3 e 5 (Projetos 7 e 6 respectivamente) devem se organizar de forma a instalar, experimentar e validar o projeto **e-gov-data-mining** disponibilizado no github em <https://github.com/MLRG-CEFET-RJ/e-gov-data-mining>. Todas as instruções para instalação e experimentação estão detalhadas no arquivo README.md na raiz do repositório. O projeto foi reorganizado e otimizado de forma a facilitar a compreensão por parte de vocês. Conforme já citado na especificação, este projeto deverá ser adaptado para servir de núcleo da ferramenta que vocês irão desenvolver.

Apesar do repositório estar funcional, ele ainda está incompleto. Em breve serão disponibilizados os recursos e funcionalidades que estão faltando. O estado atual do repositório não impede os grupos de iniciar o levantamento e análise necessários a qualquer projeto de software.

Enquanto no estado atual do repositório, os grupos devem atentar para as seguintes adaptações:

- Na aba Arquivos do canal Geral da disciplina foi criado o diretório **Datasets-Grupo3-Grupo5**. O arquivo zip que está dentro deste diretório deve ser descompactado no diretório **datasets** do projeto **e-gov-data-mining**.
- No arquivo README.md, na seção *Workflow > Run the pipeline below* vocês encontrarão a ordem em que os notebooks e scripts devem ser processados. Enquanto no estado atual do repositório vocês não devem executar o passo **RUN pre\_processamento/pre\_proc\_medicamentos.ipynb**. Este passo requer um dataset que ainda não está disponível para vocês. No entanto, o arquivo gerado a partir deste processamento está sendo disponibilizado no arquivo zip mencionado acima. Em suma, não executem esse passo por enquanto.
- Da mesma forma, o passo **RUN ean\_key\_map\_builder (not implemented)**, conforme informação explícita, não está implementado. No entanto, o arquivo gerado a partir deste processamento está sendo disponibilizado no arquivo zip mencionado anteriormente. Mais detalhes sobre este passo nas próximas reuniões presenciais de orientação.

Em caso de dúvidas ou problemas, entrem em contato pelo Teams.

Bom trabalho a todos.