**Descrição do Projeto**

1. **Introdução**

Projetos em Big Data não envolvem somente a análise dos dados e / ou apresentação de gráficos. Os dados precisam ser obtidos, armazenados, processados e partir dos mesmos obter informações relevantes.

O objetivo deste projeto é, utilizando dados de uma base publica da câmara dos deputados, obter informações relevantes sobre os gastos públicos. O aluno deverá obter os dados, armazená-los em uma base, processar eles utilizando algoritmos de machine Learning e apresentar gráficos e tabelas resumo.

1. **Desenvolvimento do Projeto Integrador 2**

Para desenvolver o seu projeto você deverá executar as seguintes fases:

* 1. **Obtenção da base de dados**

A partir do link abaixo, o aluno deverá consultar a base de gastos da Camara, no ano de 2022, montando um script que irá consultar a API e gravar os dados em um banco relacional. Podem ser utilizados bancos na máquina local OU em nuvem. O aluno deverá montar um relatório, detalhando o funcionando deste script E o modelo do banco de dados, contendo as tabelas e scripts que foram utilizados para implementação. Ao final desta etapa deve ser gerado um relatório contendo:

* + 1. **Introdução** – contextualização do trabalho proposto;
    2. **Motivação** – destacar a experiência e o problema que você quer resolver;
    3. **Script e Banco de dados –** Informações detalhadas sobre o script, tabelas que foram criadas, explicação dos campos e tipos de dados recebidos.
    4. **Considerações finais.**
  1. **Relatório Analítico**

Nesta etapa o aluno irá fazer uso dos conhecimentos adquiridos em plataforma de Bi para montar relatórios para detalhar sobre os gastos públicos, apresentando por exemplo:

* Quais deputados gastaram mais no ano analisado;
* Quais foram os principais gastos dos deputados neste ano?
* Há gastos que levantam algum tipo de suspeita?

O aluno não deverá se prender apenas as perguntas listadas e deverá montar um dashboard com informações relevantes sobre gastos públicos. Ao final o aluno deverá acrescentar no seu relatório os seguintes itens;

* + 1. **Introdução** – Explicação sobre a análise de dados publica para auxiliar a fiscalizar irregularidades.
    2. **Demonstração** – Apresentar os tipos de dados de cada campo e como eles podem ser analisados.
    3. **Relatórios e tabelas –** Gerar relatórios e tabelas contendo informações relevantes sobre a base. Incluir relatórios de linha, barra, dispersão e outros relevantes para a analise apresentada.
    4. **Considerações finais.**
  1. **Machine Learning**

Uma etapa essencial é a mineração dos dados, para identificar fatores que influenciam determinadas situações. Utilizando algoritmos de machine learning aprendidos ao longo do curso, responda no mínimo, as seguintes perguntas:

* Quais são os principais fatores que influenciaram as compras dos deputados que mais gastaram?
* Quais foram os itens / grupos de gastos com maior valor? Quais fatores informados na planilha influenciaram estes gastos?

O aluno deverá pesquisar, utilizando algoritimos de machine learning, quais fatores influenciam nos gastos dos deputados.

* + 1. **Introdução** – Detalhar sobre as técnicas de Machine Learning utilizadas
    2. **Dicionário de dados** – documentar o arquivo gerado pela captura dos dados e todos os itens de dados gerados e armazenados no seu projeto;
    3. **Considerações finais** – Destacar eventuais problemas que você enfrentou, falhas do equipamento e fatos que você julgue importante da operação e adaptações necessárias para o perfeito funcionamento do equipamento.

1. **Elaboração do relatório**

Os itens 2.1 a 2.3 devem ser desenvolvidos em um único documento, em formato ABNT. O aluno também deverá gravar um vídeo, de 10 minutos, explicando sobre todas as etapas do projeto. A nota será composta pelo relatório E pelo vídeo, então caprichem na aula.