

# Proposta do trabalho

Construção de uma inteligência artificial e processos de análise de dados para entender os dados a partir dos mesmos com a pouca interação dos clientes, com ênfase nas questões de Volume, Velocidade, Variedade dos dados gerados identificando padrões no schema e nas outras atribuições do banco de dados unificado e facilitando na hora de fazer ETL (Extrair, transformar, carregar) Validando os dados levando em conta os tipos de dados tratados e a relação entre eles ,tendo uma ferramenta de integração de dados projetada para tornar o processo de mineração de dados mais fácil e com escalonamento na hora de interpretar os dados e fazer as análises e insights do produto final.

Não esquecendo de sempre preservar a cyber segurança e backups (como se fosse um github) sem focar tanto na parte de redes complexa de ETLS com diversas fontes de dados fazendo a Gerência de dados de forma leve, consistente, adaptativa para futuras mudanças de requisitos identificando quais dados precisam ser coletados e sistematizado para que sejam visualizados de forma objetiva e pertinente.

O ideal e que esse algoritmo rode em um servidor em nuvem par evitar as grandes datas centers e diminuir com os gastos de empresas.

Concluindo Superar os desafios no uso da big data aumentando performance, armazenamento, sistematização e padronização de analises Integrando fontes diversificadas de big data possibilitando a empresa assim alinhar todo esse trabalho às adaptações e intervenções estratégicas para crescimento do empreendimento gerando insights rápidos, seguros e confiáveis.

## Pontos a serem trabalhados

- Inteligência artificial e processos de análise de dados
- Volume, Velocidade, Variedade
- Identificação de padrões
- Schema e nas outras atribuições do banco de dados unificado
- ETL (Extrair, transformar, carregar)
- Mineração de dados
- Servidor em nuvem
- Escalonamento
- Cyber segurança e backups
- Fontes de dados de redes complexas de ETL com diversas fontes de dados
- Gerência de dados de forma leve, consistente, adaptativa
- Validação os dados

## Temas e Datasets usados

- Engenharia elétrica
- Engenharia mecânica

# Ferramentas usadas

- Python
- Excel
- Power bi ou Tableau
- Spark ou IBM Datastage
- Formulários

```
9
10 from sklearn.preprocessing import LabelEncoder
11 from sklearn.model_selection import StratifiedFold
12 from sklearn.pipeline import Pipeline
13 from sklearn.model_selection import RandomizedSearchCV
14 from sklearn import model_selection
15 from sklearn.linear_model import LogisticRegression
16 from sklearn.tree import DecisionTreeClassifier
17 from sklearn.neighbors import KNeighborsClassifier
18 from sklearn.ensemble import RandomForestClassifier
19 from sklearn.naive_bayes import GaussianNB
20 from sklearn.preprocessing import StandardScaler
21 from sklearn.metrics import accuracy_score
22 from sklearn.model_selection import cross_val_predict
23 from sklearn.metrics import confusion_matrix
24 from sklearn.metrics import classification_report
25 from warnings import simplefilter
26 import pickle
27 import pandas as pd
28 import seaborn as sns
29 import matplotlib.pyplot as plt
30 from matplotlib import style
31 style.use('ggplot')
32 simplefilter(action='ignore', category=FutureWarning)
33
34
35 # Carregando e visualizando os Dados
36
37 df = pd.DataFrame()
38 for chunk in pd.read_csv('adult.data', na_values='?', chunksize=100000):
39     df = pd.concat([df, chunk])
40 df.head()
41
42 # Descrição das principais Features
43
44 # age: Idade.
45
46 # workclass: Classe trabalhadora. Se a pessoa é funcionário público, autônomo, empresário etc.
47 #
```