

[OBJETIVO DO PROJETO]

- INTRODUÇÃO BREVE.
- PROBLEMA DE NEGÓCIO.
- IMPACTO.
- MÉTRICAS DE RESULTADO.

Neste projeto cobrindo todas as etapas de um projeto real de Data Science pude resolver o problema de como utilizar dados para responder a questões importantes para permitir que uma empresa tenha conhecimento sobre:

Quais são os fatores que influenciam para um colaborador deixar a empresa?

Como reter pessoas?

Como antecipar e saber se um determinado colaborador vai sair da empresa?

E por fim disponibilizar recursos para que a empresa consiga realizar a predição para verificar se um colaborador vai ou não deixar a empresa com base em atributos como comportamento e carga de trabalho, nível de satisfação com a empresa e resultados de performance.

[SOLUÇÃO PROPOSTA]

- TECNOLOGIAS.
- ARQUITETURA.
- FRAMEWORKS.
- DESAFIOS TÉCNICOS.
- SOLUÇÕES ESCOLHIDAS.

Para resolver esse problema foi construído uma solução completa para armazenamento, gestão e automatização de fluxos de dados utilizando tecnologias como **Apache Airflow, Docker e Minio** além de explorar uma suíte poderosa de tecnologias para trabalhar com Análise de Dados e Machine Learning que são: **Pandas, Scikit-learn, Pycaret, SweetViz, Streamlit.**

Depois da infraestrutura devidamente criada e configurada, levando em consideração o desafio proposto foram criados e modelados atributos relevantes para análise utilizando fontes de dados diversas como arquivos em **formato xlsx, json** e dados no **Sistemas de Gerenciamento de Banco de Dados MySQL**.

[RESULTADOS]

- **INSIGHTS E CONHECIMENTO GERADO.**
- **PROBLEMAS RESOLVIDOS.**
- **MÉTRICAS DE PERFORMANCE.**

Na etapa de Análise Exploratória de Dados foram descobertos os vários insights importantes abaixo:

A empresa tem uma rotatividade de 24%.

Podemos assumir que os empregados que mais deixam a empresa estão menos satisfeitos.

Existe um valor considerável de empregados insatisfeitos.

A maioria dos empregados que saíram tinha salário baixo ou médio.

Os departamentos de vendas, técnico e suporte são top 3 departamentos com maior índice de turnover.

Todos os empregados que estavam inseridos em muitos projetos deixaram a empresa.

Colaboradores com baixa performance tendem a deixar a empresa.

Colaboradores insatisfeitos com a empresa têm uma maior tendência para evadir.

Através da análise foi possível desenvolver 3 grupos distintos para agrupar colaboradores que deixaram a empresa por comportamentos similares que são:

Grupo 1 (Empregados insatisfeitos e trabalhadores): A satisfação foi inferior a 20 e as avaliações foram superiores a 75.

Que corresponde ao grupo de funcionários que deixaram a empresa e eram bons trabalhadores, mas se sentiam péssimos no trabalho.

Grupo 2 (Empregados ruins e insatisfeitos): Satisfação entre 35 à 50 e as suas avaliações abaixo de ~ 58.

Corresponde aos empregados que foram mal avaliados e se sentiram mal no trabalho.

Grupo 3 (Empregados satisfeitos e trabalhadores):

Representa os empregados ideais, que gostam do seu trabalho e são bem avaliados por seu desempenho.

Este grupo pode indicar os empregados que deixaram a empresa porque encontraram outra oportunidade de trabalho.

Para a estimativa com o objetivo de prever se um empregado vai deixar a empresa foi implementado um modelo utilizando o algoritmo **Gradient Boosting Classifier** que atingiu uma performance de **AUC em 0.80**.

[CONCLUSÃO]

- **RESUMO FINAL DO PROJETO.**
- **BENEFÍCIOS ALCANÇADOS.**
- **POSSÍVEIS MELHORIAS.**

Através desse projeto foi possível praticar e implementar conceitos importantes da Ciência e Engenharia de Dados e propor uma solução para um problema latente e recorrente de qualquer empresa que é a retenção de talentos através da Análise de Dados de Recursos Humanos.

Como um processo de melhoria contínua podemos desenvolver uma automação para executar não só o pipeline de coleta e transformação de dados como automatizar os passos da etapa de Machine Learning e Deploy.