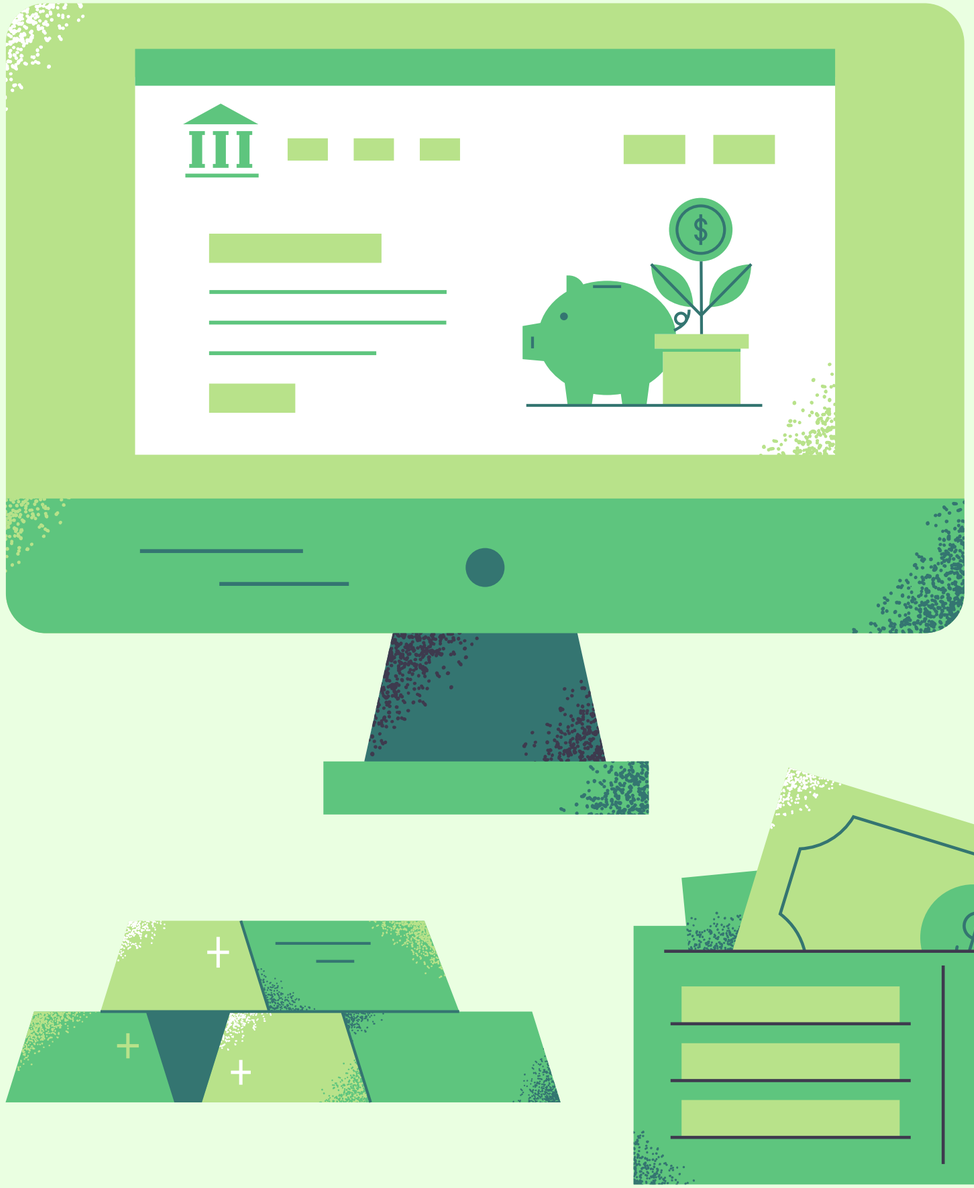
ANALISE SALARIAL

PROJECTO FINAL



CONTEXTO DE ESTUDO

Num mercado de trabalho cada vez mais competitivo, compreender os fatores que influenciam os salários é essencial para promover justiça interna, atratividade de talento e eficiência organizacional.

Compreender os fatores que influenciam os salários é essencial para promover justiça interna e atrair talento num mercado de trabalho competitivo. Esta análise baseia-se em dados fiáveis, devidamente tratados e normalizados, contamos por isso com:

+35k	15	24,1%
Registos	variáveis	Salários +50k

ROADMAP DA NOSSA IMPLEMENTAÇÃO

Como objectivo pretende-se desenvolver um projecto em contexto real

Diagnosticar

Os fatores mais importantes na remoneração

Desenvolver

Modelos preditivos explicáveis

Segmentar

Nossos dados com métodos não supervisionados

Descobrir

Padrões e relações através de regras de associação

Dashboard

Disponibilizar dashboard apra consulta em tempo real

CONTENT STRATEGY ROADMAP

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus lacinia odio vitae vestibulum. Nullam quis risus eget urna mollis ornare vel eu leo.

Estrutura de dados

Compreender e organizar os dados disponíveis, garantindo uma base sólida para a análise futura.

Limpeza e Preparação

Eliminar ruído, inconsistências e valores inválidos, assegurando a qualidade e integridade dos dados.

Análise Exploratória

Aplicar técnicas exploratórias e estatísticas para extrair padrões relevantes e gerar insights acionáveis.

Modelação Supervisionada

Aplicação de algoritmos como Random Forest e Regressão Logística para previsão de salários com validação cruzada e análise de desempenho.

Clusterização e Segmentação

Utilização de K-Means e DBSCAN para identificar perfis ocultos de trabalhadores com base em variáveis socio-económica

CONTENT STRATEGY ROADMAP

Regras de Associação

Descoberta de padrões salariais frequentes com Apriori probabilidade de >50K), gerando 62.599 regras úteis para análise estratégica.

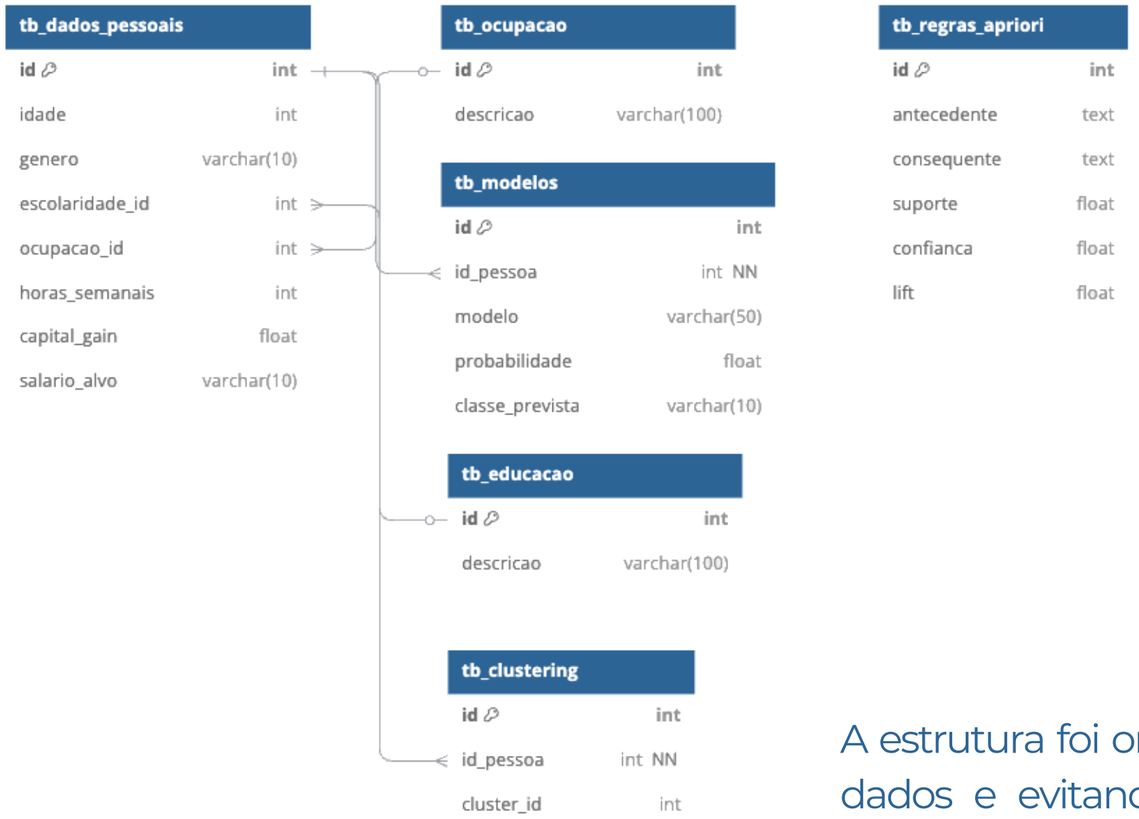
Visualização e Comunicação

Desenvolvimento de dashboards interativos com Streamlit e criação de views SQL para apoio à decisão e reporting automático

Ajustes e Validações Finais

Avaliação de limitações, validação ética (sem viés discriminatório) proposta de melhorias com integração de SMOTE, novos modelos (XGBoost)

ESTRUTURA DE BASE DADOS



TABELAS
Total

6

TABELAS
Relacionadas

5

RESULTADOS
Analiticos

3

A estrutura foi organizada em modelo relacional, garantindo a integridade dos dados e evitando redundâncias. Foram criadas tabelas específicas para os registos principais e categorização de variáveis, facilitando queries analíticas e construção de views especializadas.

Analise de dados

Age	Idade
Workclass	Classe profissional
fnlwgt	Peso amostral
education-tatus	Escolaridade
marital-status	Estado Civil
ocupation	Tipo de profissão
relationship	Relação com o agregado familiar
race	Grupo Étnico
sex	Sexo
capital-gain	Ganho de Capital
capital-loss	Perda de Capital
hours-per-week	Carga Horária semanal
native-country	País de Nascimento
salary	Faixa Salárial - <=50K e > 50k

PRINCIPAIS RESULTADOS

Carga Horária

Verifica-se que a maioria dos individuos tem uma carga horaria de **35 - 40 Horas semanais**

Idade Média

A idade média dos individuos situa-se entre **38 anos**

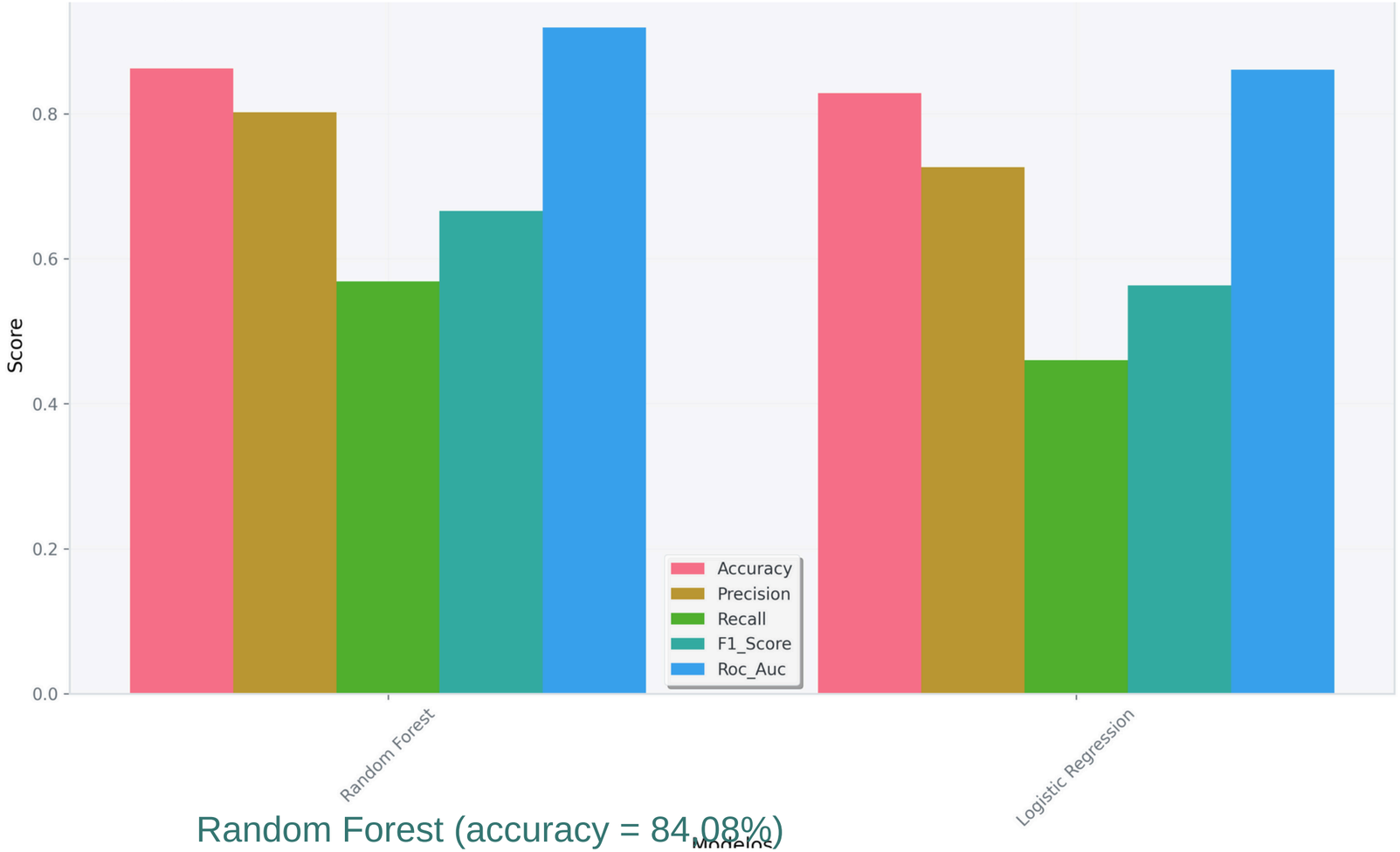
Sexo

Verifica-se que os indivuduos do sexo masculino.
Mais de **30%** dos Individuos com ganham mais de 50k são do **sexo masculino**.

Escolaridade

Verifica-se que **+55%** dos individuos que ganham **+50k** são titulares do grau académico **Mestrado**

COMPARAÇÃO DE PERFORMANCE DE MODELOS

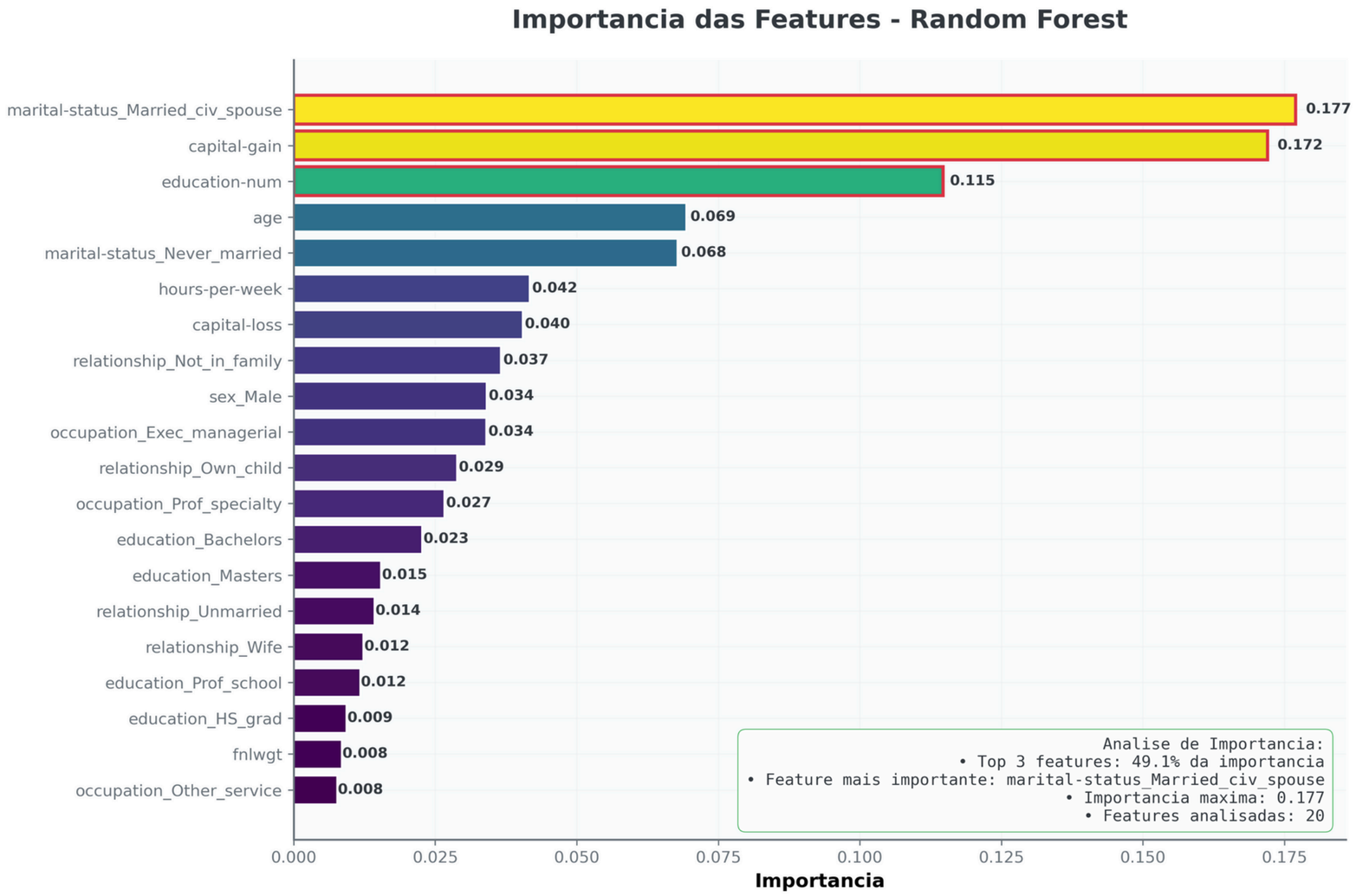


Random Forest (accuracy = 84,08%)

- Regressão Logística (accuracy = 81,85%)
- Métricas: Precision, Recall, F1-Score, Matriz de Confusão
- Validaçãocruzada (K-Fold = 5)
- Interpretação dos coeficientes e importância das features

RANDOM FOREST

Modelo supervisionado



Top 3 Features

Casados
17,7% Decisão
Possibilidade de viés sociais

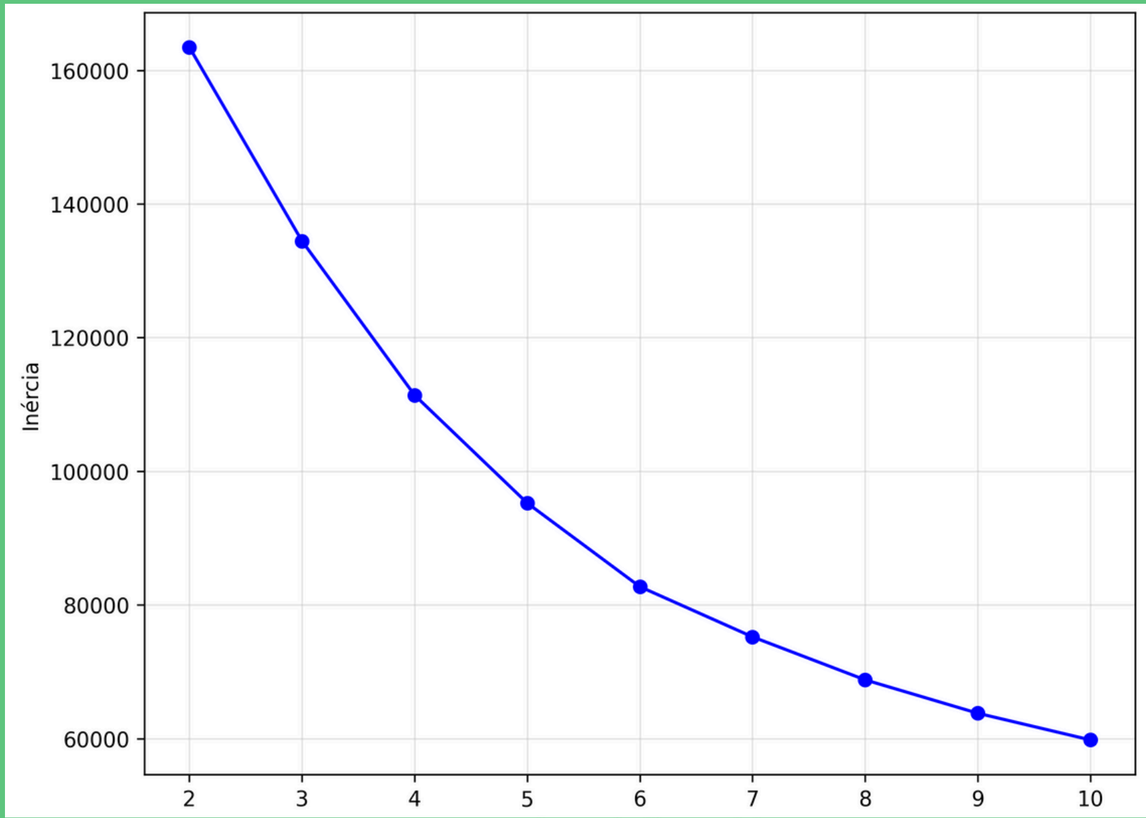
Ganhos de capital
17,2% Decisão
Possibilidade rendimento extra

Grau Académico
11,5% Decisão

46,4% Importância total

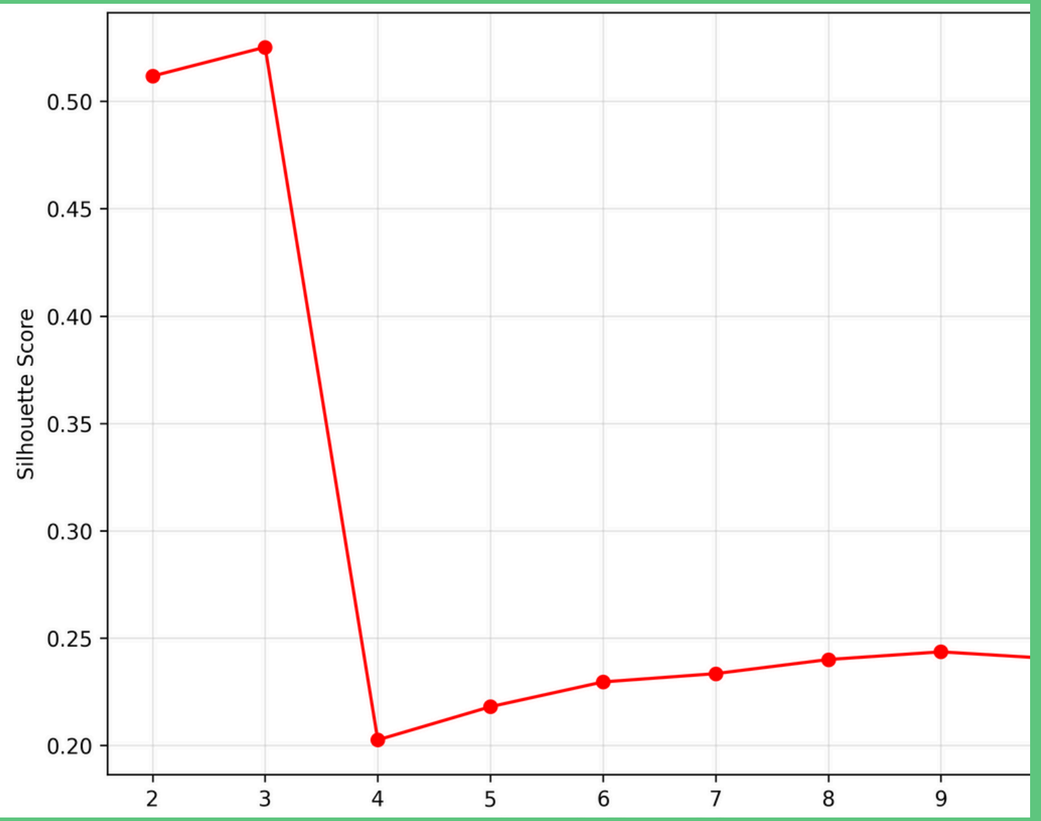
A análise de importância das features pelo Random Forest indica que fatores como estado civil, ganhos de capital e escolaridade têm elevado poder preditivo para distinguir entre salários elevados e reduzidos, independentemente de apresentarem correlações lineares com outras variáveis.

K-MEANS



“COTOVELO”

Redução acentuada da inércia ao aumentar K de 2 até 3 ou 4
O “cotovelo” do gráfico ocorre por volta de K=3 ou K=4.
Neste caso, K=3 parece um valor razoável, pois, a partir deste valor, a redução da inércia é mais suave.

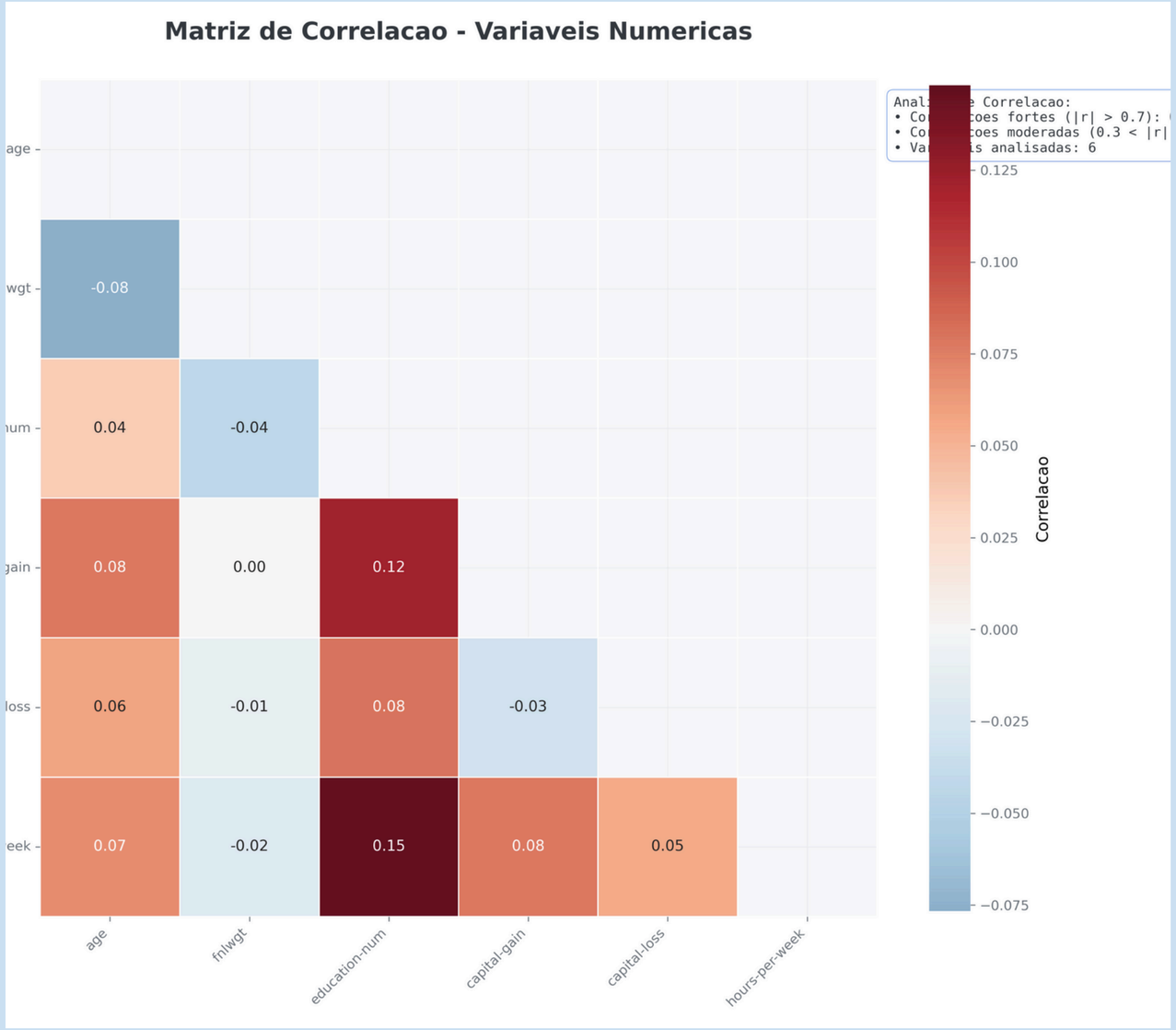


SILHOUETTE SCORE

Os valores mais altos para K=2 e K=3, indicando clusters bem definidos e separado

MATRIX DE CORRELAÇÃO

- Não há correlações fortes nem moderadas entre as variáveis analisadas.
- Não existem correlações fortes (|r| > 0.7) nem moderadas (0.3 < |r| < 0.7) neste conjunto.
- Todos os coeficientes estão entre -0.08 e 0.15, ou seja, muito próximos de zero.
- O maior valor é education-num vs hours-per-week com r = 0.15 (fraco positivo).
- capital-gain tem uma fraca correlação positiva com education-num (r = 0.12).
- age tem uma correlação ligeiramente negativa com fnlwgt (r = -0.08).



As relações identificadas pelos modelos de machine learning não devem ser interpretadas como causais, mas sim como preditivas. A análise cruzada entre correlação e importância das features permite maior robustez e transparência na interpretação dos resultados.

CONCLUSÃO

01 LIMITAÇÕES DA ANÁLISE:

Apesar destas evidências, as correlações lineares entre variáveis numéricas revelaram-se fracas, o que limita a interpretação causal direta. Além disso, a amostra pode apresentar vieses de género, escolaridade ou ocupação, não totalmente controlados neste estudo.

02 FORÇA DA ABORDAGEM PREDITIVA

Os modelos de machine learning demonstraram maior capacidade para identificar padrões relevantes na previsão do salário, captando relações não lineares e interações complexas entre variáveis, ultrapassando as limitações da estatística descritiva tradicional.

03 SUGESTÕES PARA TRABALHO FUTURO

Recomenda-se aprofundar a análise com técnicas de explicabilidade (ex.: SHAP), explorar novos atributos ou fontes de dados e desenvolver abordagens para mitigação de vieses. Uma análise mais fina dos segmentos de menor representatividade (por exemplo, mulheres com salários elevados) também seria pertinente.

04 REFLEXÃO ÉTICA E CRÍTICA:

É essencial garantir que as soluções desenvolvidas promovam equidade, transparência e responsabilidade, nomeadamente face ao potencial impacto de decisões algorítmicas em processos de gestão salarial e oportunidades de carreira

ESTÁS PRREPARADO PARA CONHECER OS NUMEROS DA TUA EMPRESA?

E-mail dariodourado@gmail.com

Website www.dariodouradodev.com

Phone +351-937-372-716

OBRIGADO

DARIO DOURADO