

Análise de Dados: Diabetes e Fatores Correlatos

Sprint 03 – Data Science

FIAP

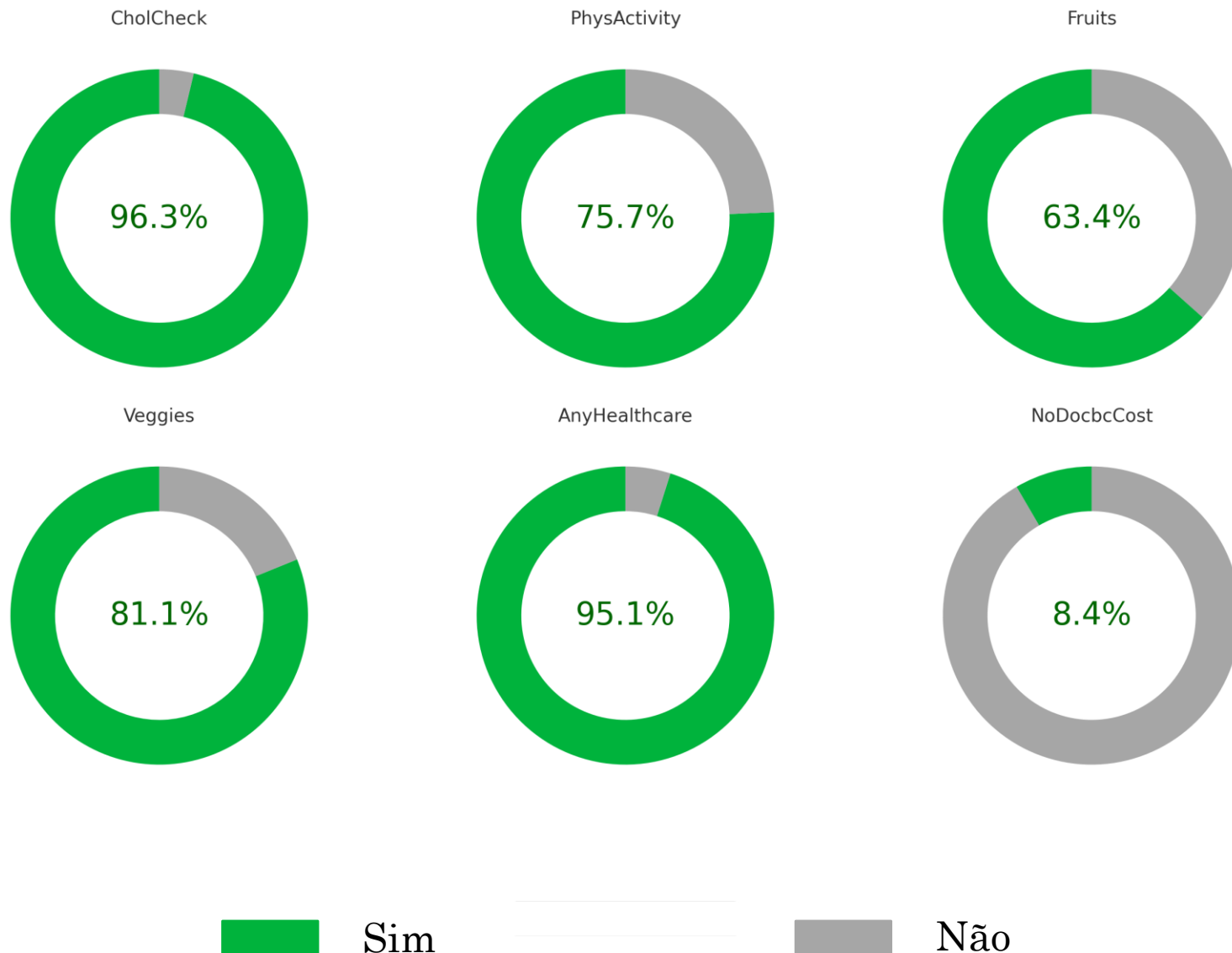
Sumário

Objetivo

Analisar o conjunto de dados para identificar os fatores de risco associados ao diabetes, explorando o inter-relacionamento das variáveis e sua correlação com a presença da doença.

Resumo Executivo

A análise revela que variáveis como IMC, pressão arterial e atividade física têm correlação significativa com o risco de diabetes. O modelo preditivo criado para estimar o status de diabetes não atingiu os 80% de precisão esperados, ficando em torno de 74%.



Estatística Descritiva

Essas informações nos dão uma ideia do perfil de saúde e hábitos da população estudada.

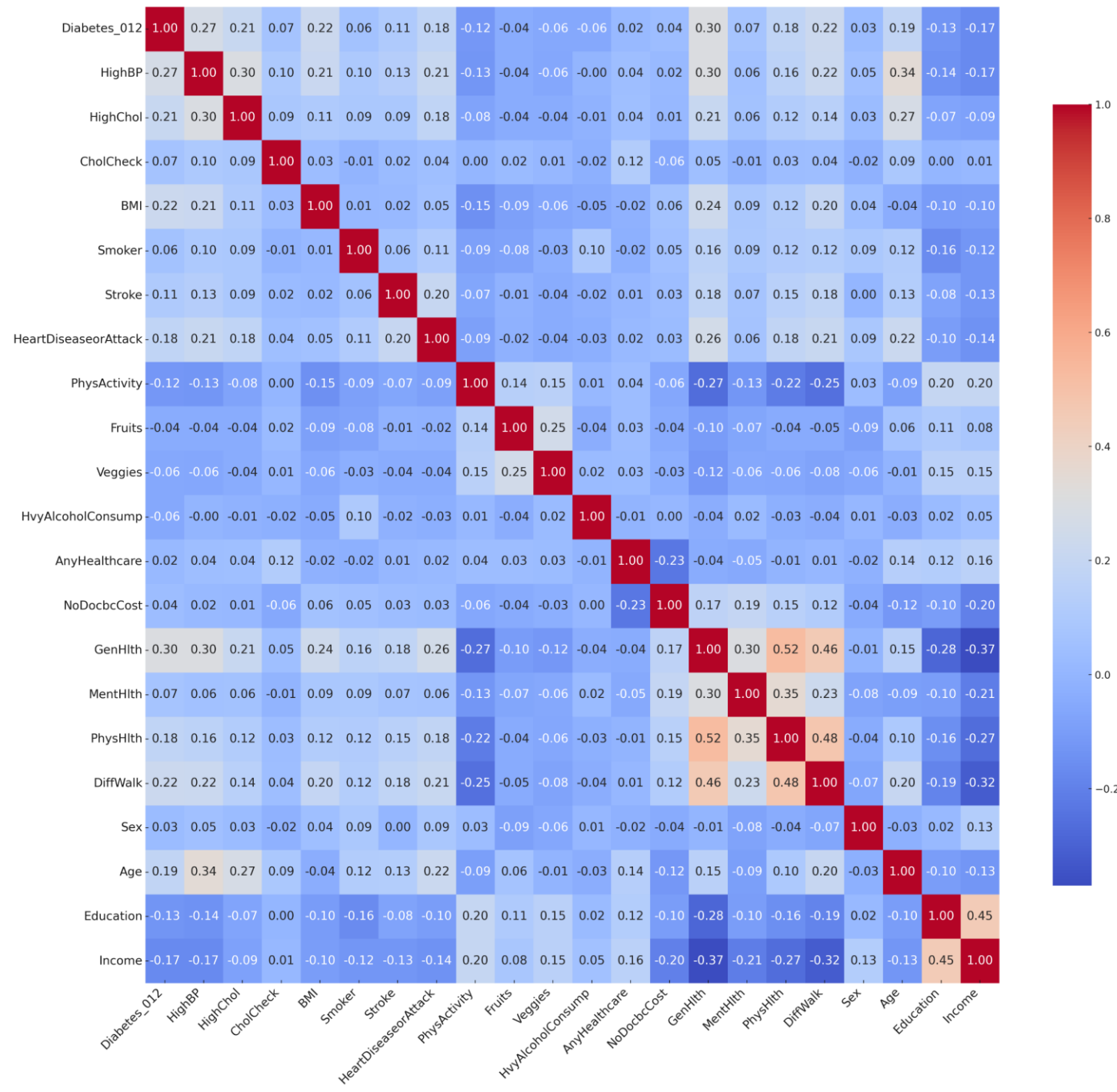
Matriz de correlação

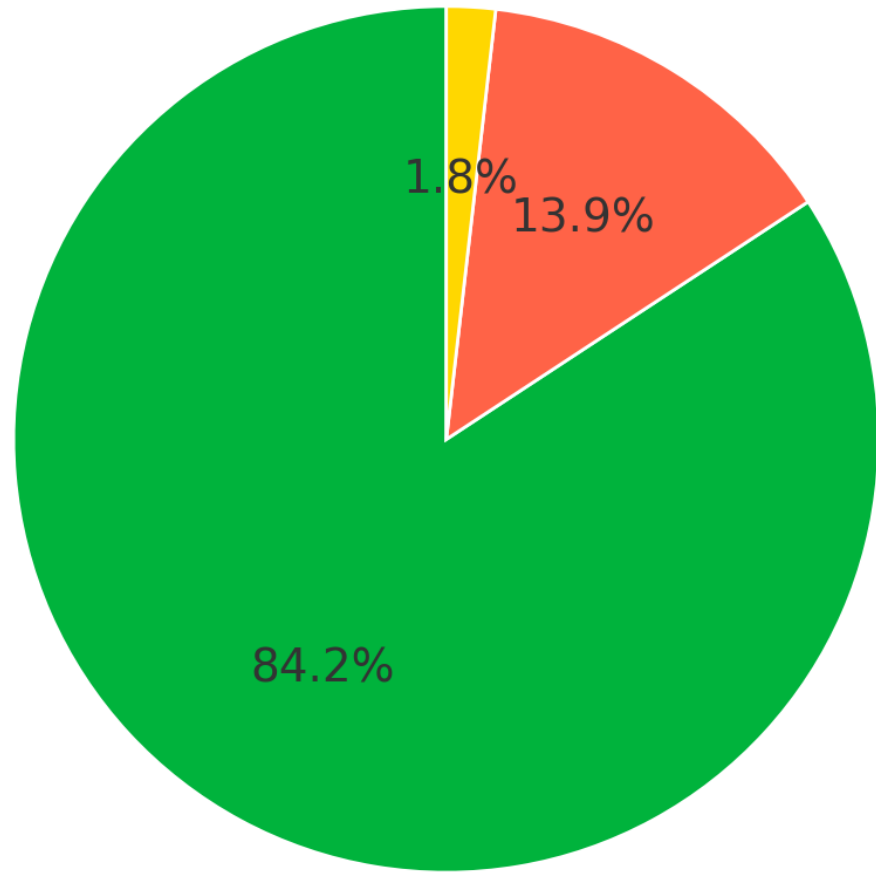
Diabetes: Correlação positiva com pressão alta (0.27), colesterol alto (0.30) e IMC elevado (0.22).

Atividade Física: Reduz o risco de diabetes (-0.12), pressão alta (-0.13) e doenças cardíacas (-0.13).

Saúde Mental: Níveis socioeconômicos mais baixos estão associados a pior saúde mental (renda: -0.21, educação: -0.28).

IMC: Elevado IMC está relacionado à pressão alta (0.21) e menor atividade física (-0.15).





74% de taxa de acerto.

O valor esperado de **80%** poderia ter sido alcançado caso a distribuição dos valores de **Diabetes**, **Pré-diabetes** e **Sem Diabetes** fosse mais uniforme.

Comparado a base tratada, a base crua apresentou uma pequena melhora de **2%** no percentual de acerto, mas ainda abaixo do esperado.

Teste de Hipóteses: Modelo de Classificação

Considerações Finais

O estudo realizado demonstrou que o banco de dados é adequado para entender as correlações entre as variáveis de saúde e o risco de diabetes. Variáveis como IMC, pressão alta e nível de atividade física mostraram-se significativamente associadas à presença de diabetes, corroborando o conhecimento médico existente. Além disso, o modelo preditivo desenvolvido obteve uma precisão moderada, ainda que abaixo dos 80% esperados, sugerindo que o conjunto de dados pode ser melhorado para aumentar a eficácia preditiva, especialmente ao lidar com o desbalanceamento de classes.

A remoção de outliers não proporcionou um aumento na precisão do modelo, destacando que o tratamento desses valores deve ser cuidadosamente considerado, dependendo do impacto nas análises subsequentes. No geral, o projeto cumpre seu objetivo de explorar as inter-relações entre as variáveis e oferece uma base sólida para o desenvolvimento de modelos preditivos mais robustos, especialmente com ajustes nas classes minoritárias e variáveis-chave.

Grupo

- Caíque Walter Silva - RM550693
- Gabriela Marsiglia - RM551237
- Guilherme Nobre Bernardo - RM98604
- Matheus José de Lima Costa - RM551157