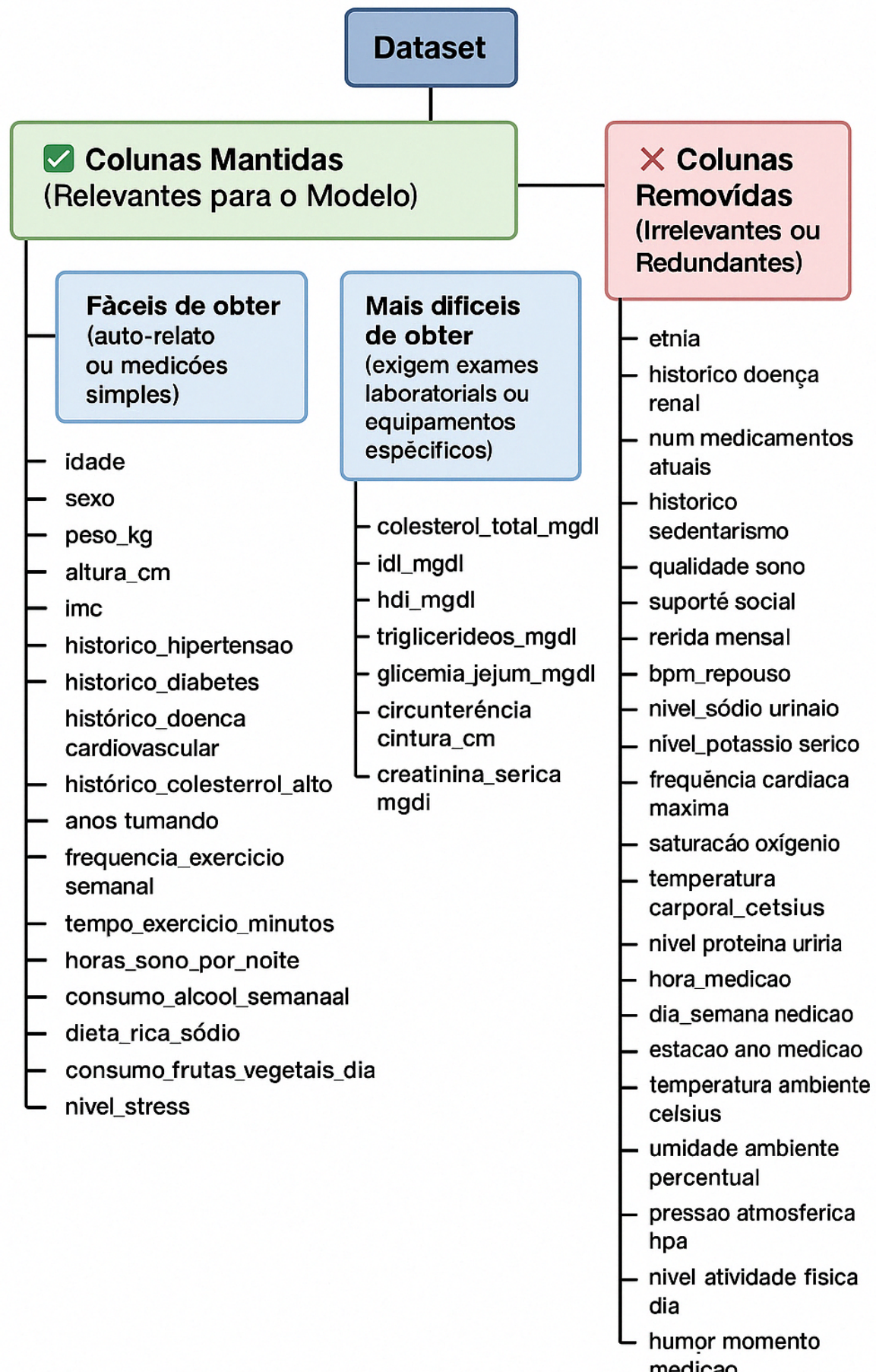


Introdução ao Design do Dataset

Este documento descreve a estruturação de um dataset voltado para o treinamento de um sistema de inteligência artificial com o objetivo de auxiliar pacientes na identificação da propensão à hipertensão.

As variáveis foram classificadas conforme sua relevância clínica e acessibilidade prática. Foram mantidos apenas os dados com forte correlação com a condição estudada. As variáveis mantidas foram ainda divididas em subgrupos conforme a facilidade de obtenção dos dados, e as variáveis removidas foram excluídas por apresentarem características redundantes, subjetivas, ou irrelevantes clinicamente.



Justificativas para Remoção de Variáveis

etnia:

Pode introduzir viés e não é diretamente associada ao risco individual de hipertensão.

historico_doenca_renal:

Menos diretamente relacionada do que outros fatores clínicos selecionados.

num_medicamentos_atuais:

Não diferencia entre tipos de medicamentos, gerando ruído.

historico_sedentarismo:

Redundante, já há frequência e tempo de exercício.

qualidade_sono:

Já representado por horas de sono por noite.

suporte_social:

Subjetivo, difícil de medir com precisão.

renda_mensal:

Pode estar correlacionado a outras variáveis, mas é indireto.

bpm_repouso:

Pode ser consequência da hipertensão, e não causa.

nivel_sodio_urinario:

Requer exame difícil e é menos acessível.

nivel_potassio_serico:

Laboratorial, menos relevante frente a outras variáveis mais preditivas.

frequencia_cardiaca_maxima:

Mais útil em contexto de esforço físico e não cotidiano.

saturacao_oxigenio:

Mais voltada a problemas pulmonares do que cardiovasculares.

temperatura_corporal_celsius:

Não influencia diretamente a hipertensão.

nivel_proteina_urina:

Indicador de problemas renais mais que de pressão.

hora_medicao:

Variação natural diária não impacta a tendência do indivíduo.

dia_semana_medicao:

Sem relação causal com hipertensão.

estacao_ano_medicao:

Pode afetar pressão pontualmente, mas não tendência crônica.

temperatura_ambiente_celsius:

Afeta medições pontuais, não tendência geral.

umidade_ambiente_percentual:

Mesma justificativa da temperatura.

pressao_atmosferica_hpa:

Sem evidência de impacto relevante na tendência.

nivel_atividade_fisica_dia:

Coberto por frequência/tempo de exercício.

qualidade_alimentacao_dia:

Dados mais objetivos já foram mantidos.

humor_momento_medicao:

Extremamente subjetivo e volátil.

Organização do Dataset de Pressão Arterial

Variáveis Mantidas		Variáveis Removidas
Fáceis de obter	Mais difíceis	
idade	colesterol_total_mgdl	etnia
sexo	ldl_mgdl	historico_doenca_renal
peso_kg	hdl_mgdl	num_medicamentos_atuais
altura_cm	triglicerideos_mgdl	historico_sedentarismo
imc	glicemia_jejum_mgdl	qualidade_sono
historico_hipertensao	circunferencia_cintura_cm	suporte_social
historico_diabetes	creatinina_serica_mgdl	renda_mensal
historico_doenca_cardiova		bpm_repouso
historico_colesterol_alto		nivel_sodio_urinario
historico_tabagismo		nivel_potassio_serico
anos_fumando		frequencia_cardiaca_maxima
frequencia_exercicio_sema		saturacao_oxigenio
tempo_exercicio_minutos		temperatura_corporal_celsius
horas_sono_por_noite		nivel_proteina_urina
consumo_alcool_semanal		hora_medicao
dieta_rica_sodio		dia_semana_medicao
consumo_frutas_vegetais_		estacao_ano_medicao
nivel_stress		temperatura_ambiente_celsius
		umidade_ambiente_percentual
		pressao_atmosferica_hpa
		nivel_atividade_fisica_dia
		qualidade_alimentacao_dia
		humor_momento_medicao
Variável Alvo: label_pressao_alta		