



## **Sprint Review 3:** Otimização da rede neural com protótipo de alta fidelidade

# Agenda

01. | **Sprint 2**  
Visualização das entregas e sua timeline

02. | **Entregas anteriores**  
Recapitulação das entregas passadas

03. | **Otimização e hiperparâmetros**  
Melhoria do modelo de rede neural

04. | **Resultados**  
Performance da rede neural otimizada

05. | **Protótipo**  
Construção do protótipo navegável de alta fidelidade

06. | **Próximos passos**  
Visão geral do planejamento



# Sprint 3

Melhoria do modelo de rede neural junto ao protótipo de alta fidelidade da plataforma



SPRINT 1

Análise exploratória dos dados, Canvas Proposta de Valor, Matriz de Risco, Personas e Anti personas e Jornada do Usuário



SPRINT 2

User Stories, Wireframes e Modelo de Rede Neural



SPRINT 3

**Otimização do modelo, protótipo de alta fidelidade, Análise PESTEL e Business Model Canvas**



SPRINT 4

Avaliação das Métricas e Front End inicial

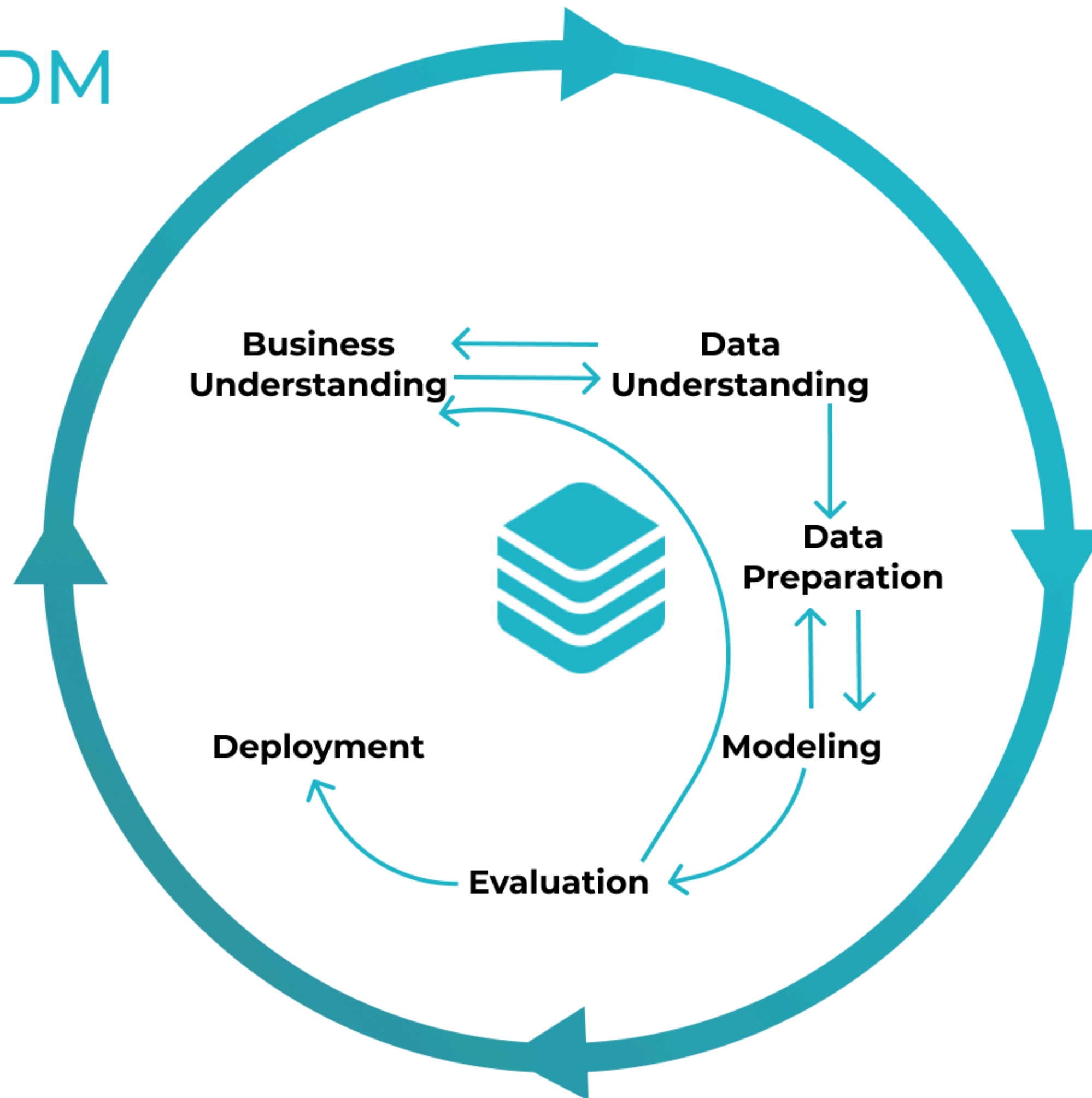


SPRINT 5

Disponibilização do modelo, front end final e Análise Financeira do projeto



# CRISP-DM







# Pipeline - Sprint 2



Transformação do dataset para parquet

Parquet fará com que haja menor ocupação de espaço e memória



Remoção de colunas que não serão usadas

Ex: EMP\_CODIGO / ECO\_OUTRAS / FATURADO\_MEDIA, etc.



Conversão dos tipos de dados

Para inteiros (exceto lat e long) e datetime



Normalização dos dados numéricos

Escala dos dados entre 0 e 1, usando RobustScaler



Padronização do texto

Conversão para maiúsculas e remoção de caracteres especiais



One hot encoding

Representação numérica de variáveis categorias



Criação da categoria de consumo

Consumo alto  $> 0.7$   
 $0.2 \leq$  Consumo médio  $\leq 0.7$   
Consumo baixo  $< 0.2$



Redimensionamento do Dataframe

1 matrícula por linha e colunas para cada mês de consumo, de cada ano



Adição da coluna Target - Fraude

1 matrícula por linha e colunas para cada mês de consumo, de cada ano



# Pipeline - Sprint 3



Aplicação do cuDF



Remoção de colunas que não serão usadas

Ex: Grande quantidade de colunas



Conversão dos tipos de dados

Para inteiros (exceto lat e long) e datetime



Normalização dos dados numéricos

Escala dos dados entre 0 e 1, usando RobustScaler



Padronização do texto

Conversão para maiúsculas e remoção de caracteres especiais



One hot encoding

Representação numérica de variáveis categorias



Criação da categoria de consumo

Consumo alto  $> 0.7$   
 $0.2 \leq$  Consumo médio  $\leq 0.7$   
Consumo baixo  $< 0.2$



Redimensionamento do Dataframe

1 matrícula por linha e colunas para cada mês de consumo, de cada ano



Adição da coluna Target - Fraude

1 matrícula por linha e colunas para cada mês de consumo, de cada ano



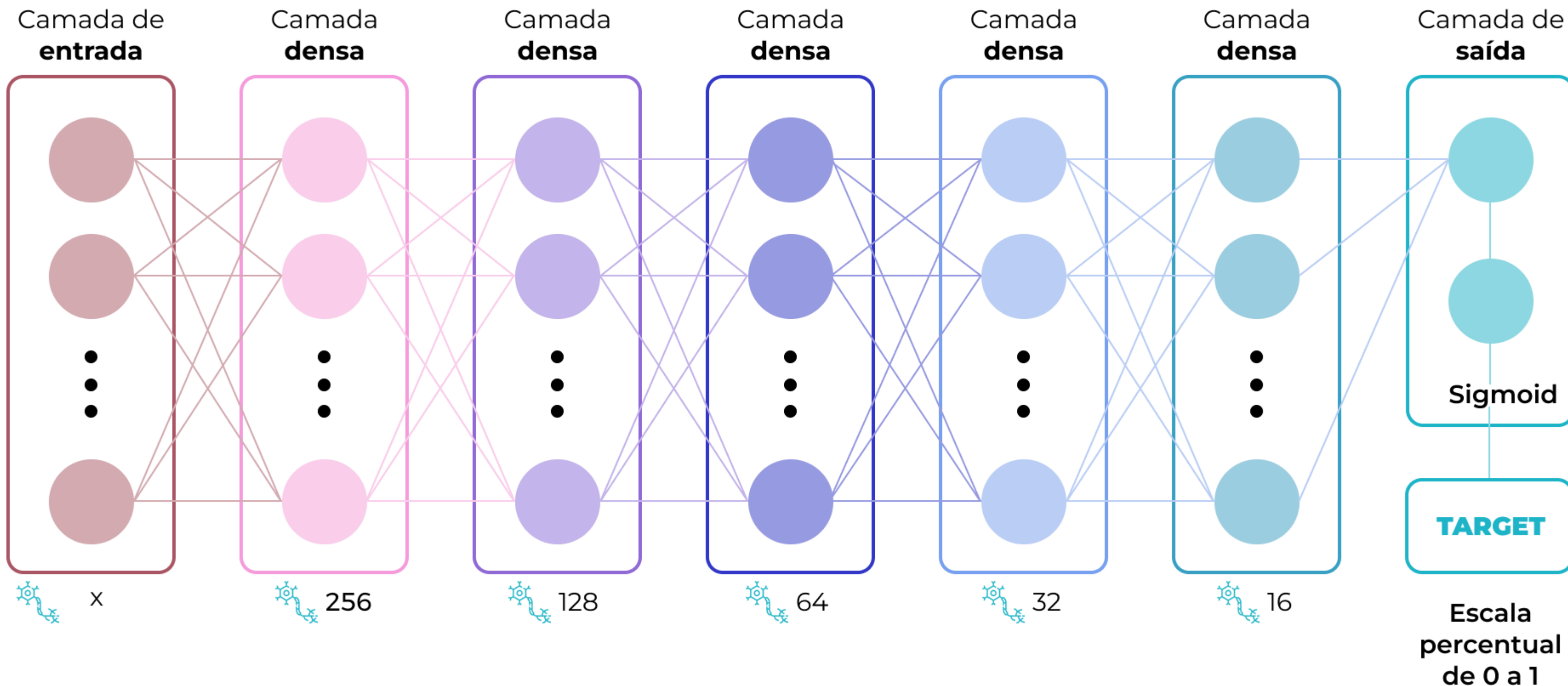
UnderSample e dados de input ao modelo

Float e int para o modelo +  
Mais UnderSample dos dados



# Arquitetura - Sprint 2

Feedforward (FNN)

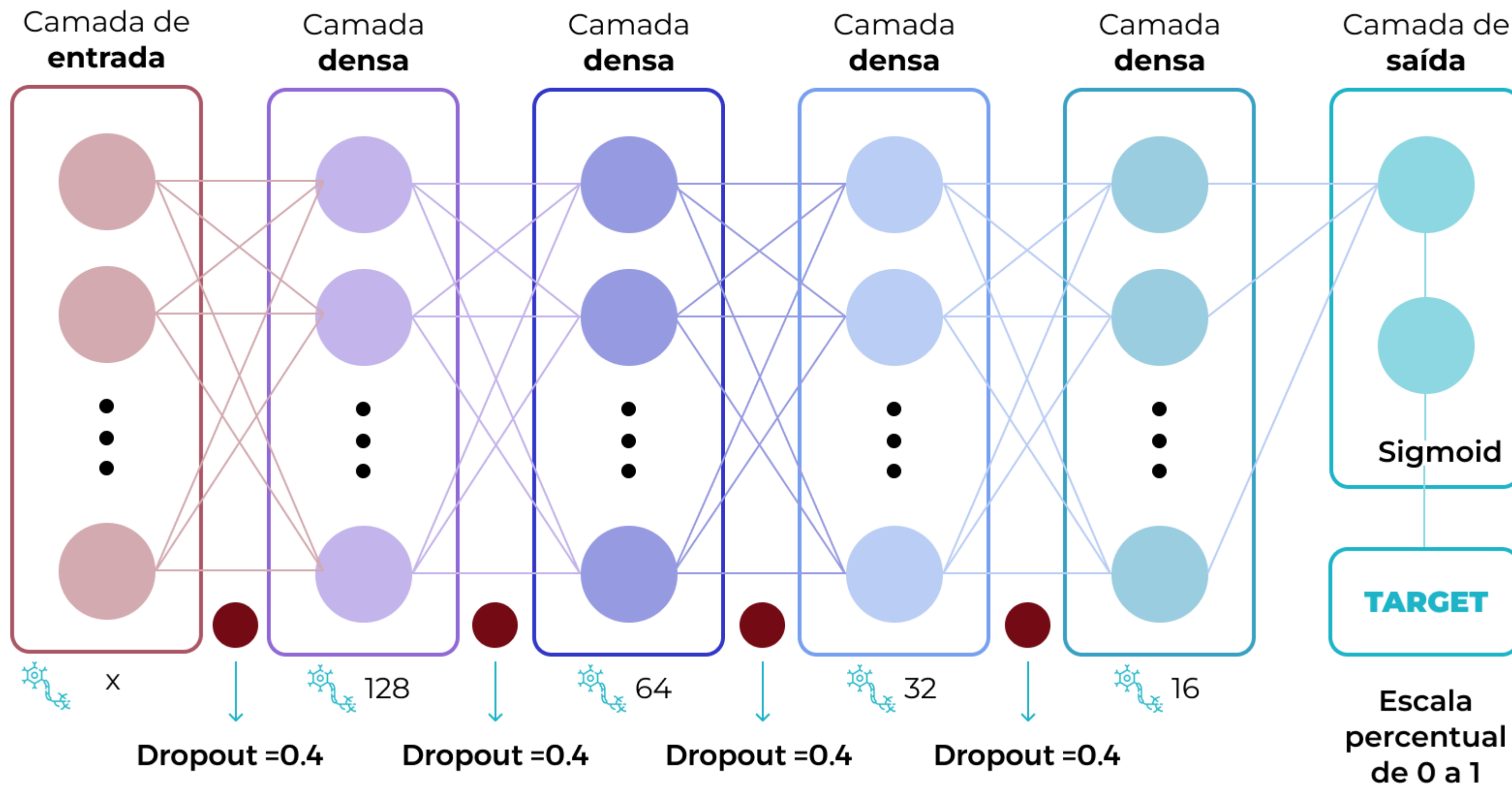






# Arquitetura - Sprint 3

Feedforward (FNN)







# Otimização e hiperparâmetros



Random search: busca a melhor configuração da rede testando combinações aleatórias de parâmetros

**1. melhor otimizador**

**2. melhor taxa de aprendizado**

**3. melhor quantidade de épocas**

**4. melhor tamanho do batch size**



ajusta pesos para detectar  
melhor as fraudes

define o ritmo para o ajuste e o  
modelo aprender  
eficientemente

número ideal de vezes que os  
dados serão revisitados

exemplos de dados que processa  
antes de ajustar os pesos



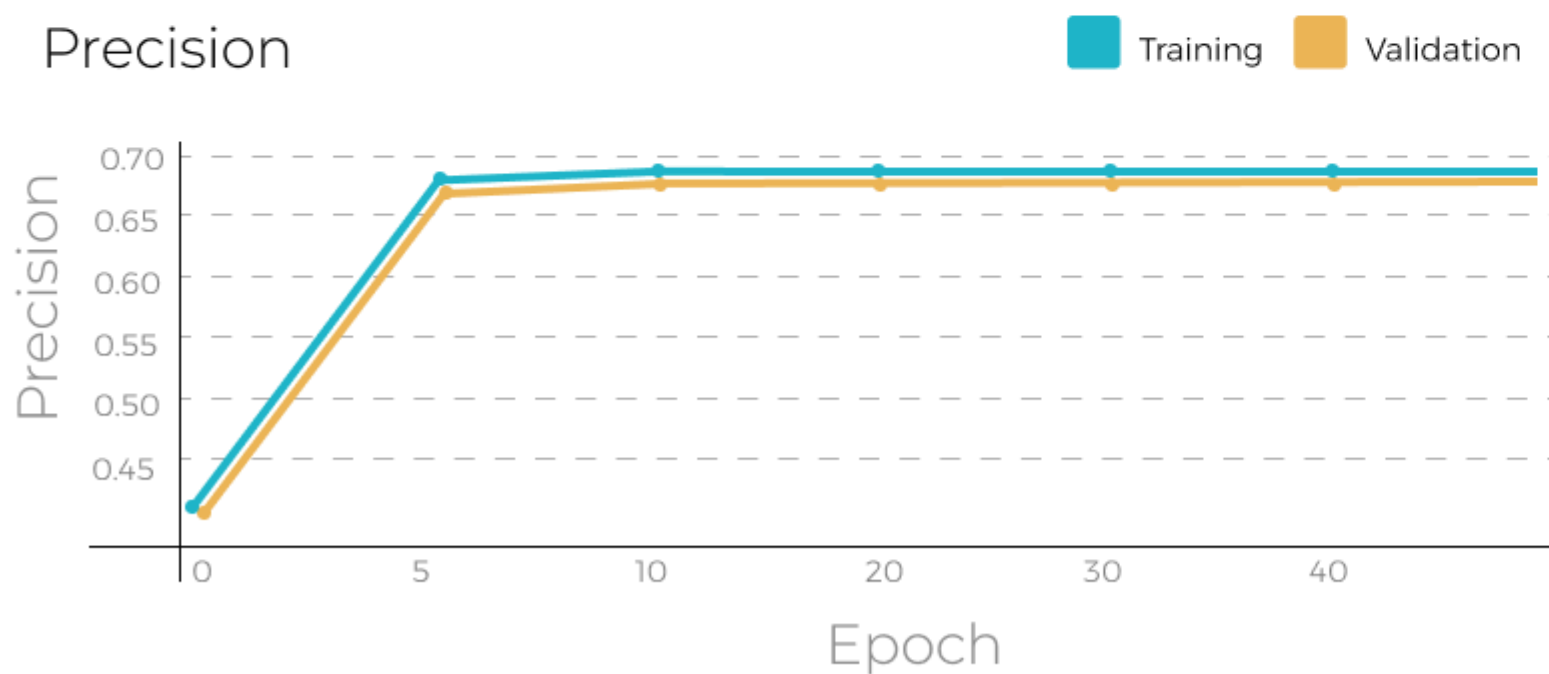
# Resultados

## *precisão*

Avalia o percentual de predições corretas de fraudes e não fraudes entre todas as predições feitas

Treino: 0.68

Teste: 0.67

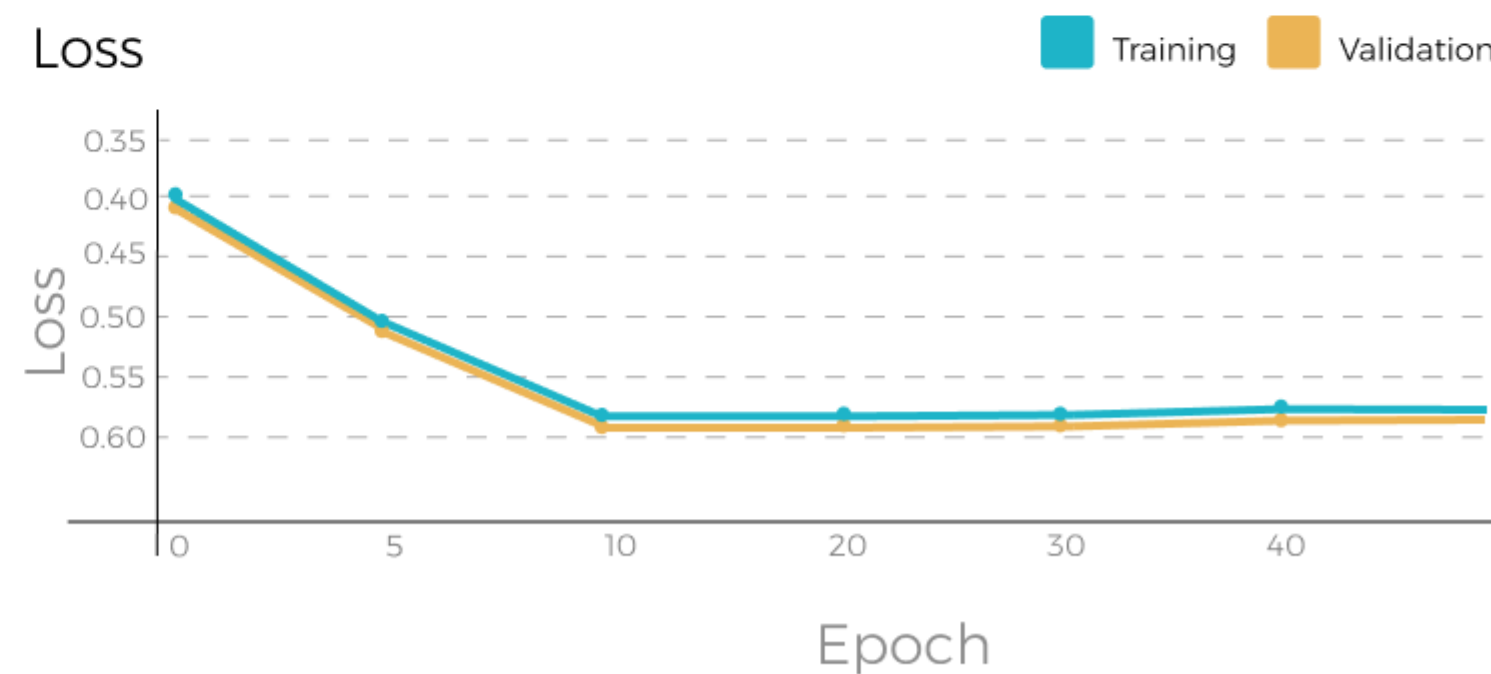


## *loss*

Avalia o quão bem o modelo está se ajustando aos dados de treino, medindo a discrepância entre as predições e os valores reais

Treino: 0.58

Teste: 0.57





# Protótipo





# Próximos passos



01

## RELATÓRIO DE AVALIAÇÃO DAS MÉTRICAS

análise dos erros, curva de aprendizado



02

## CAPACIDADE DE GENERALIZAÇÃO DO MODELO

com base nos resultados obtidos no conjunto em teste



03

## PRIMEIRA VERSÃO DO FRONT END

implementação inicial no Streamlit



Estamos abertos  
a responder  
quaisquer  
dúvidas



Sophia Dias



Vitória  
Rodrigues



Pedro  
Rezende



Camila  
Anacleto



Raduan  
Muarrek



Patrick  
Miranda



Henri  
Harari

