**🤖 Página 6: Modelagem Preditiva e Detecção de Fraude**

**📌 Visão Geral da Modelagem**

Nesta fase do projeto, aplicamos técnicas avançadas de aprendizado de máquina para desenvolver modelos capazes de prever e detectar automaticamente casos suspeitos de fraudes nas entregas do Walmart. Utilizamos algoritmos poderosos como Random Forest, Gradient Boosting e Redes Neurais.

Todas as implementações foram feitas em Python utilizando bibliotecas como Scikit-learn, XGBoost e TensorFlow.

**🛠️ Técnicas e Algoritmos Utilizados**

**1. Random Forest Classifier**

Utilizamos o Random Forest para prever a probabilidade de fraude em cada pedido:

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report, confusion\_matrix

# Preparação dos dados

X = orders\_df[['order\_amount', 'items\_delivered', 'delivery\_hour']]

y = orders\_df['items\_missing'].apply(lambda x: 1 if x > 0 else 0)

# Dividindo o dataset

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Treinamento do modelo

rf\_model = RandomForestClassifier(n\_estimators=100, random\_state=42)

rf\_model.fit(X\_train, y\_train)

# Avaliação do modelo

predictions = rf\_model.predict(X\_test)

print(classification\_report(y\_test, predictions))

sns.heatmap(confusion\_matrix(y\_test, predictions), annot=True, fmt='d', cmap='Blues')

plt.title('Matriz de Confusão - Random Forest')

plt.xlabel('Predito')

plt.ylabel('Real')

plt.show()

**2. Gradient Boosting (XGBoost)**

Usamos o XGBoost para melhorar a capacidade preditiva e reduzir falsos positivos:

import xgboost as xgb

# Treinamento do modelo XGBoost

xgb\_model = xgb.XGBClassifier(n\_estimators=100, learning\_rate=0.05, random\_state=42)

xgb\_model.fit(X\_train, y\_train)

# Avaliação do modelo

xgb\_predictions = xgb\_model.predict(X\_test)

print(classification\_report(y\_test, xgb\_predictions))

sns.heatmap(confusion\_matrix(y\_test, xgb\_predictions), annot=True, fmt='d', cmap='Greens')

plt.title('Matriz de Confusão - XGBoost')

plt.xlabel('Predito')

plt.ylabel('Real')

plt.show()

**3. Rede Neural (TensorFlow)**

Implementamos uma rede neural para capturar padrões mais complexos nos dados:

import tensorflow as tf

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

# Construindo a rede neural

nn\_model = Sequential([

Dense(64, activation='relu', input\_shape=(X\_train.shape[1],)),

Dense(32, activation='relu'),

Dense(1, activation='sigmoid')

])

# Compilando o modelo

nn\_model.compile(optimizer='adam', loss='binary\_crossentropy', metrics=['accuracy'])

# Treinando o modelo

nn\_model.fit(X\_train, y\_train, epochs=50, batch\_size=32, validation\_split=0.2)

# Avaliação do modelo

nn\_predictions = (nn\_model.predict(X\_test) > 0.5).astype(int)

print(classification\_report(y\_test, nn\_predictions))

sns.heatmap(confusion\_matrix(y\_test, nn\_predictions), annot=True, fmt='d', cmap='Reds')

plt.title('Matriz de Confusão - Rede Neural')

plt.xlabel('Predito')

plt.ylabel('Real')

plt.show()

**📊 Avaliação e Seleção do Melhor Modelo**

Após treinar e avaliar os modelos, realizamos uma comparação detalhada para selecionar o modelo com melhor desempenho em termos de precisão, recall e f1-score.

* Random Forest e XGBoost demonstraram excelente performance e facilidade interpretativa.
* Redes Neurais mostraram potencial para capturar padrões complexos, mas requerem mais tempo e recursos.

**Critérios para Seleção:**

* Melhor balanceamento entre precisão e recall.
* Capacidade de generalização para novos dados.

**🎯 Principais Insights e Decisões**

* Modelo selecionado (ex: XGBoost) será implementado para monitoramento contínuo.
* Implementação prática com alertas automáticos para facilitar a tomada de decisão rápida.

**🚀 Próximo Passo: Avaliação das Responsabilidades**

Com um modelo preditivo robusto desenvolvido, avançaremos agora para a investigação detalhada das responsabilidades e causas associadas às fraudes detectadas, oferecendo insights práticos para intervenções e medidas preventivas eficazes.