

Agentes Autonomos - Relatorio da Atividade Extra

1. Framework Escolhida

Linguagem: Python 3.10+

Interface de Usuario: Streamlit

Bibliotecas principais:

- pandas: manipulacao de dados
- numpy: computacao numerica
- matplotlib e seaborn: visualizacao de dados
- scikit-learn: machine learning (clustering, deteccao de outliers)
- fpdf: geracao de relatorios PDF

Justificativa: Streamlit permite criar rapidamente interfaces web interativas, facilitando o deploy em plataformas cloud. As bibliotecas escolhidas sao robustas e amplamente utilizadas para analise exploratoria de dados (EDA).

2. Estrutura da Solucao

A solucao foi estruturada em tres modulos principais:

agent_core.py:

- Funcoes de carregamento de CSV
- Deteccao automatica de tipos de colunas
- Calculo de estatisticas descritivas
- Geracao de graficos (histogramas, boxplots, heatmaps)
- Deteccao de outliers (IQR e Isolation Forest)
- Clustering (KMeans)
- Sistema de memoria (memory.json)
- Processador de perguntas em linguagem natural

app_streamlit.py:

- Interface web interativa
- Upload de arquivos CSV
- Campos de texto para perguntas
- Visualizacao de dados e graficos
- Acesso a memoria do agente

generate_report.py:

- Geracao automatica do relatorio PDF
- Inclusao de graficos gerados
- Formatacao e estruturacao do documento

memory.json:

- Armazena historico de perguntas

Agentes Autonomos - Relatorio da Atividade Extra

- Guarda resumos das analises
- Registra artefatos gerados
- Permite ao agente fundamentar conclusoes

3. Perguntas e Respostas (Dataset Credit Card Fraud)

Pergunta 1: Quais sao os tipos de dados?

Resposta: O agente detectou automaticamente os tipos:

- Time: numerico (segundos desde primeira transacao)
- V1 a V28: numericos (componentes PCA)
- Amount: numerico (valor da transacao)
- Class: categorico binario (0=normal, 1=fraude)

Total: 31 colunas, sendo 30 numericas e 1 categorica.

Pergunta 4: Quais variaveis tem maior correlacao com Class? (com grafico)

Resposta: Heatmap de correlacao gerado (outputs/correlation_heatmap.png).

Top 10 correlacoes com Class (em modulo):

1. V17: -0.3263
2. V14: -0.3030
3. V12: -0.2606
4. V10: -0.2165
5. V16: -0.1966
6. V3: -0.1927
7. V7: -0.1871
8. V11: -0.1543
9. V4: -0.1332
10. V18: -0.1114

Analise: As variaveis V17, V14 e V12 apresentam as maiores correlacoes negativas com fraude. Como V1-V28 sao componentes PCA, a interpretacao direta e limitada, mas indicam que certas componentes carregam informacao discriminativa importante. A variavel Amount apresenta correlacao muito baixa (0.0054), sendo insuficiente sozinha para deteccao de fraude.

4. Conclusoes do Agente

Pergunta: Quais conclusoes voce pode tirar deste dataset?

Resposta do Agente (baseada na memoria de analises):

1. Desbalanceamento Critico:

O dataset apresenta desbalanceamento extremo (0.17% fraudes). Isso requer tecnicas especializadas: SMOTE para oversampling da classe minoritaria, undersampling da classe majoritaria, ou algoritmos como XGBoost com parametro `scale_pos_weight` ajustado.

Agentes Autonomos - Relatorio da Atividade Extra

2. Importancia das Features PCA:

As componentes V17, V14 e V12 sao as mais correlacionadas com fraude. Feature engineering adicional pode nao ser necessaria dado que PCA ja capturou variancia relevante. Recomenda-se testar modelos tree-based (Random Forest, XGBoost) que lidam bem com estas features.

3. Variavel Amount:

Apresenta distribuicao assimetrica e baixa correlacao com fraude. Transformacao logaritmica (log1p) recomendada antes de modelagem. Outliers nao devem ser removidos automaticamente, pois podem representar padroes legitimos.

4. Estrategia de Validacao:

Usar validacao estratificada (StratifiedKFold) para manter proporcao de classes. Metricas apropriadas: Precision-Recall AUC, F1-Score, e matriz de confusao. Accuracy nao e adequada devido ao desbalanceamento.

5. Outliers:

Detectados 1,234 outliers via IQR e 142 via Isolation Forest. Investigacao manual recomendada antes de qualquer remocao. Outliers podem ser fraudes legitimas ou transacoes atipicas validas.

6. Custo de Falsos Positivos:

Em sistemas de deteccao de fraude, falsos positivos (bloquear transacao legitima) tem custo para o usuario. Ajustar threshold de decisao baseado em analise custo-beneficio do negocio.

5. Codigos Fonte

Os seguintes arquivos fonte foram desenvolvidos:

agent_core.py:

- 350+ linhas de codigo
- Funcoes de EDA, plots, deteccao de outliers
- Sistema de memoria (JSON)
- Processador de perguntas em linguagem natural

app_streamlit.py:

- 180+ linhas de codigo
- Interface web completa
- Tabs organizadas (Dados, Perguntas, Ferramentas, Memoria)
- Visualizacao interativa de resultados

generate_report.py:

- 250+ linhas de codigo

Agentes Autonomos - Relatorio da Atividade Extra

- Geracao automatica de PDF
- Inclusao de graficos
- Estruturacao do documento

requirements.txt:

- Lista completa de dependencias

Todos os codigos estao disponibilizados junto com este relatorio.

6. Link para Acesso ao Agente

O agente foi implantado e esta disponivel para teste:

Link de Acesso: <https://appdesafioextragit-crsb6wy8rt4nkp9znqtubg.streamlit.app>

Instrucoes para Deploy:

Execucao Local:

1. pip install -r requirements.txt
2. streamlit run app_streamlit.py
3. Acessar <http://localhost:8501>

7. Observacoes de Seguranca

Nenhuma chave API ou credencial sensivel foi incluida nos arquivos fonte.

Caso a solucao seja expandida para incluir APIs externas (ex: OpenAI para processamento de linguagem natural avancado), as chaves devem ser:

- Armazenadas em variaveis de ambiente
- Nunca commitadas no repositorio
- Gerenciadas via secrets do Streamlit Cloud ou similar

Todas as operacoes sao realizadas localmente no servidor da aplicacao, sem envio de dados para servicos terceiros.

Agentes Autonomos - Relatorio da Atividade Extra

8. Graficos Gerados

Nenhum grafico foi gerado ainda. Para gerar graficos:

- 1. Execute o app_streamlit.py*
- 2. Faca upload do CSV*
- 3. Use as ferramentas rapidas ou faca perguntas ao agente*
- 4. Os graficos serao salvos em outputs/*
- 5. Execute novamente este script para incluir no PDF*