



## ESPECIFICAÇÕES DO PROJETO

**Identify Fraud from Enron Email****Qualidade do Código**

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Funcionalidade	O código reflete a descrição presente nas respostas das perguntas no relatório. O código faz as funções documentadas no relatório e o relatório especifica claramente a estratégia de análise final.
Usabilidade	poi_id.py pode ser rodado para exportar o dataset, lista de features e algoritmo, de maneira que o algoritmo final pode ser verificado usando o tester.py.

**Entendimento dos Dados e da Pergunta**

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Exploração dos Dados	A resposta do estudante trata as características mais importantes do conjunto de dados e usa estas

CRITÉRIO	características para fazer as suas análises. Características importantes incluem: <b>ATENDEU ÀS ESPECIFICAÇÕES</b>
	<ul style="list-style-type: none"> <li>• número total de data points</li> <li>• alocação entre classes (POI/non-POI)</li> <li>• número de características usadas</li> <li>• existem características com muitos valores faltando? etc.</li> </ul> <p>Dica: você pode utilizar visualizações para auxiliar nesse processo</p>
Investigação dos Outliers	O aluno identifica o(s) outlier(s) nos dados financeiros e explica como eles foram removidos ou tratados

### Otimização da Seleção de Características/Engenharia

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Criação de novas features	Pelo menos uma feature foi implementada. A justificativa para ela foi dada nas respostas escritas e o efeito desta característica na performance final foi testada. O aluno não precisa incluir a nova característica no conjunto de dados final.
Seleção de features feita de forma inteligente	Seleção de características univariadas ou recursivas foi feita ou as características foram escolhidas manualmente (diferentes combinações de características foram feitas e o desempenho foi documentado para cada uma delas). Método de seleção e características selecionadas são documentadas e o número selecionado foi justificado. Para um algoritmo que suporta a verificação da importância das variáveis (ex. decision tree) ou pontuação das características (ex. SelectKBest), estas estão documentadas também.

<b>CRITÉRIO</b> Ajuste de escala das características feito corretamente	<b>ATENDEU ÀS ESPECIFICAÇÕES</b> Se o algoritmo requerir características com ajuste de escala, esta foi feita nos dados.
--	---

### Escolha e Ajustes de um Algoritmo

<b>CRITÉRIO</b>	<b>ATENDEU ÀS ESPECIFICAÇÕES</b>
Escolha um algoritmo	Pelo menos 2 algoritmos diferentes são usados e seus desempenhos são comparados, com o de melhor desempenho sendo usado no modelo final. Essa comparação deve ser devidamente reportada no relatório
Discussão da afinação de parâmetros e a performance deles.	A resposta endereça o que significa fazer o afinamento(tuning) dos parâmetros e porque é importante fazê-lo.
Afinação (tuning) do algoritmo	Pelo menos um parâmetro importante com 3 valores é investigado sistematicamente, ou qualquer dos seguintes são verdadeiros:

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
	<ul style="list-style-type: none"> <li>GridSearchCV usado para a busca do melhor parâmetro</li> <li>Vários parâmetros são afinados</li> <li>Busca de parâmetros incorporado na seleção do algoritmo (ex. parâmetros afinados para mais de um algoritmo e a melhor combinação algoritmo-parâmetros selecionada para a análise final).</li> </ul>

### Validar e Avaliar

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Uso das Métricas de Avaliação	Pelo menos duas métricas apropriadas são usadas para avaliar a performance do algoritmo (ex: precisão e abrangência - precision and recall), e o aluno explica o que estas métricas medem no contexto desta tarefa.
Discussão de validação e a sua importância.	A resposta explica o que é validação e porque ela é importante.
Estratégia de Validação	O desempenho do modelo final é medido dividindo a base de dados entre base de treinamento e teste ou através do uso de validação cruzada (cross validation), especificando o tipo de validação usado.

Performance do Modelo CRITÉRIO	Quando tester.py é usado para avaliar a performance, precision e recall são, os dois, ao menos 0.3. ATENDEU AS ESPECIFICAÇÕES
--------------------------------------	---

---