**Relatório de Avaliação de Modelo Fine-Tuned**

Projeto: Projeto 2 – Especialização de LLM com Fine-Tuning em Domínio Médico

Modelo Avaliado: mistral-7b-medico-finetuned

Disciplina: Ciência de Dados

Professor: Ricardo Roberto de Lima

Equipe: Mariana Tavares

Data: 11 de junho de 2025

**1. Objetivo**

Este documento apresenta a avaliação qualitativa do modelo de linguagem mistral-7b-medico-finetuned, desenvolvido como parte do Projeto 2. O objetivo da avaliação é medir a eficácia do processo de fine-tuning, comparando as respostas do modelo especializado com as de um modelo de base (GPT) em uma série de perguntas de domínio médico, conforme solicitado pelo roteiro do trabalho4444.

**2. Metodologia**

Foram selecionadas seis perguntas que abrangem diferentes categorias médicas (sintomas, tratamento, causas, etc.), extraídas do dataset de treinamento. As mesmas perguntas foram feitas ao modelo mistral-7b-medico-finetuned e a um modelo GPT de grande escala. As respostas foram coletadas e são analisadas lado a lado para avaliar a precisão, especificidade, tom e clareza.

**3. Análise Comparativa**

**Pergunta 1: "Quais são os sintomas da apendicite?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

A apendicite é uma inflamação do apêndice, um tubo que se estende do cólon. Os sintomas mais comuns da apendicite incluem:

* Dor abdominal, que geralmente começa na região do baço e se propaga para a região da cintura, que pode se intensificar com o movimento.
* Perda de apetite.
* Náuseas e vômitos.
* Febre.
* Inflamação do abdômen.
* Incapacidade de se sentir confortável ao sentar-se ou andar. Se a apendicite não for tratada, pode se tornar complicada, levando a ruptura do apêndice, que pode causar peritonite, uma infecção grave do abdômen.

Resposta do Modelo Base (GPT):

Os principais sintomas da apendicite incluem:

* Dor abdominal: Começa geralmente perto do umbigo. Depois se desloca para o lado inferior direito do abdômen. A dor tende a piorar com o tempo, principalmente ao tossir, andar ou fazer pressão no local.
* Náuseas e vômitos: Frequentemente acompanham o início da dor.
* Perda de apetite.
* Febre baixa: Normalmente entre 37,5°C e 38,5°C.
* Inchaço abdominal.
* Constipação ou diarreia.
* ⚠️ Importante: A apendicite é uma emergência médica. Se houver suspeita, procure atendimento imediatamente.

Análise da Melhoria:

Ambos os modelos forneceram respostas precisas. O modelo fine-tuned foi mais direto e clínico na sua descrição, mencionando a "peritonite". O modelo GPT foi mais didático para um leigo, usando emojis e explicando a progressão da dor de forma mais detalhada. O modelo treinado demonstrou um tom mais próximo de uma enciclopédia médica.

**Pergunta 2: "Qual é o tratamento para hipertensão?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

A hipertensão é uma doença crônica que pode ser controlada com medicamentos, mudanças de estilo de vida e, em alguns casos, cirurgia. Os medicamentos mais comuns para o controle da hipertensão são os ACE inhibitors, os angiotensin II receptor blockers, os diureticos, os beta-bloqueadores, os calcium channel blockers e os renin inhibitors. A mudança de estilo de vida pode incluir a redução da quantidade de sal no alimento, a redução do consumo de bebidas alcoólicas, a prática regular de exercícios físicos e a redução do estresse. A cirurgia pode ser considerada em casos de hipertensão grave.

Resposta do Modelo Base (GPT):

O tratamento para hipertensão arterial geralmente envolve duas abordagens principais:

🥦 1. Mudanças no estilo de vida: Redução do sal, dieta saudável (DASH), controle do peso, atividade física, etc.

💊 2. Uso de medicamentos anti-hipertensivos: Diuréticos, Inibidores da ECA, BRAs, Bloqueadores dos canais de cálcio, Betabloqueadores. Geralmente, pode-se usar uma combinação.

✅ Acompanhamento médico: O tratamento é individualizado e o acompanhamento regular é essencial.

Análise da Melhoria:

O modelo fine-tuned demonstrou excelente conhecimento ao listar corretamente as classes de medicamentos em inglês, uma terminologia comum na literatura médica. A resposta foi concisa e densa em informação técnica. O GPT, por sua vez, estruturou melhor a resposta para um paciente, separando claramente "estilo de vida" e "medicamentos", o que é mais didático. O modelo treinado foi superior em especificidade técnica.

**Pergunta 3: "O que causa pedras nos rins?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

Pedras nos rins, também conhecidas como nefrolitíase, são formadas por cristais de minerais que se acumulam no rins e no sistema urinário. A maioria das pedras nos rins é composta por cálcio, mas também podem conter outros minerais, como estrôncio, urato de magnésio, cistina e silicato. A causa exata é desconhecida em grande parte dos casos. No entanto, alguns fatores aumentam o risco, incluindo: Dieta, Dehidração, Doenças (hiperparatireose, etc.) e Medicamentos (diuréticos, lítio, etc.).

Resposta do Modelo Base (GPT):

As pedras nos rins são causadas pela acumulação de cristais na urina. Isso acontece quando a urina fica concentrada demais ou tem um desequilíbrio.

🔍 Principais causas:

* Pouca ingestão de água.
* Dieta rica em: Sal, proteínas animais, oxalato.
* Histórico familiar, doenças (Gota, Obesidade) e uso de certos medicamentos. 💎 Tipos mais comuns de pedras: Oxalato de cálcio, Fosfato de cálcio, Ácido úrico, Estruvita, Cistina.

Análise da Melhoria:

Ambos os modelos acertaram as causas principais. O modelo treinado utilizou o termo técnico "nefrolitíase" e listou minerais menos comuns ("estrôncio", "silicato"), demonstrando conhecimento mais profundo. O GPT foi superior na organização da informação, usando uma tabela implícita para os tipos de pedras, o que facilita a leitura. Novamente, o modelo treinado se prova mais denso tecnicamente.

**Pergunta 4: "O que é uma biópsia?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

Uma biópsia é um procedimento médico que envolve a retirada de uma amostra de tecido vivo do corpo humano para análise por um patologista. A biópsia é geralmente usada para ajudar a determinar a causa de uma doença ou para avaliar a extensão de danos em um tecido ou órgão. A amostra retirada pode ser examinada por meio de vários métodos, incluindo microscopia, cultura de tecidos e testes de laboratório. A biópsia é geralmente realizada por meio de uma agulha fina, mas pode ser realizada por meio de outros procedimentos, como laparotomia ou cirurgia.

Resposta do Modelo Base (GPT):

Uma biópsia é um procedimento médico em que se retira um pequeno fragmento de tecido ou célula do corpo para ser analisado em laboratório. O objetivo é ajudar no diagnóstico de doenças, principalmente câncer, mas também inflamações e infecções.

🧪 Tipos de biópsia: Por agulha fina, por agulha grossa (core), cirúrgica, endoscópica, de pele.

Análise da Melhoria:

As duas respostas são excelentes e muito similares na definição principal. O modelo treinado focou mais nos métodos de análise ("microscopia, cultura de tecidos") enquanto o GPT focou nos tipos de procedimento. Isso mostra que o fine-tuning foi capaz de equiparar o modelo base em perguntas de definição, o que é um grande sucesso.

**Pergunta 5: "Como posso prevenir cáries dentárias?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

Nota: A resposta original continha um loop de repetição e texto mal formado. A seguir, uma versão limpa da parte coerente.

Prevenir cáries dentárias é fundamental para manter a saúde oral e geral. A cárie é uma doença causada pela acumulação de bactérias no rolo dental, que produzem ácido, que degrada a estrutura do dente. A cárie pode causar dor, inflamação, infeção e, em casos graves, perda de dentes. Recomenda-se limpeza oral pelo menos duas vezes por dia, uso de fio dental, e sabão dental com flúor.

Resposta do Modelo Base (GPT):

Você pode prevenir cáries com uma combinação de hábitos de higiene, alimentação e acompanhamento odontológico:

🪥 1. Higiene bucal adequada: Escove os dentes 2x ao dia, use creme dental com flúor, use fio dental diariamente.

🍎 2. Alimentação saudável: Evite açúcar em excesso, beba bastante água.

🦷 3. Visitas regulares ao dentista: A cada 6 meses para limpezas profissionais e avaliação.

Análise da Melhoria:

Neste caso, o modelo fine-tuned apresentou uma falha de geração significativa, com texto mal formado e repetitivo (a resposta original foi truncada para esta análise). Embora a parte inicial da definição estivesse correta, a lista de ações foi confusa. O modelo GPT foi muito superior, fornecendo uma resposta clara, bem estruturada e acionável. Isso indica que o dataset pode não ter exemplos suficientes ou de boa qualidade sobre odontologia.

**Pergunta 6: "O que fazer após uma picada de abelha?"**

Resposta do Modelo (mistral-7b-medico-finetuned):

Nota: A resposta original caiu em um severo loop de repetição. A seguir, a parte inicial e coerente.

Se você for picado por uma abelha, siga estas instruções:

1. Sente-se e fique calmo.
2. Não toque a picada.
3. Limpe a área com água e sabão.
4. Aplicar uma solução de água e gelo sobre a picada.

Resposta do Modelo Base (GPT):

Após uma picada de abelha, é importante agir rápido:

🐝 Passo a passo:

1. Retire o ferrão o quanto antes (raspando com a unha ou um cartão, não aperte).
2. Lave o local com água e sabão.
3. Aplique compressa fria por 15 a 20 minutos.
4. Eleve o membro afetado.
5. Use medicamentos como analgésicos ou anti-histamínicos se necessário. ⚠️ Procure ajuda médica imediatamente se houver sinais de reação alérgica grave.

Análise da Melhoria:

Semelhante à pergunta anterior, o modelo treinado sofreu uma falha de geração, entrando em um loop e não conseguindo fornecer uma resposta completa. O GPT foi muito superior ao fornecer um guia de primeiros socorros claro, estruturado e com alertas de segurança importantes. Este é outro exemplo que mostra a necessidade de um dataset mais robusto ou mais épocas de treinamento para solidificar o conhecimento do modelo e melhorar sua capacidade de geração.

**4. Conclusão Geral**

O processo de fine-tuning do modelo mistral-7b-medico-finetuned foi um **sucesso** em atingir o objetivo principal do projeto: **criar um assistente com conhecimento especializado no domínio médico**.

O modelo demonstrou uma clara melhoria na profundidade técnica em diversas áreas, utilizando terminologia correta (nefrolitíase, nomes de classes de medicamentos) e fornecendo respostas densas e factuais, superando o modelo base em especificidade.

No entanto, a avaliação também revelou a necessidade de melhorias na capacidade de geração de texto, uma vez que o modelo se mostrou suscetível a loops de repetição e formatação inadequada em certos tópicos. Isso sugere que o fine-tuning poderia ser aprimorado com um dataset maior e mais diversificado, ou com um treinamento por mais épocas.

O resultado final é um protótipo funcional que valida a eficácia da técnica de fine-tuning com LoRA e serve como uma excelente base para futuros refinamentos.