



EXTRAÇÃO DE HbA1C DE PRONTUÁRIOS MÉDICOS UTILIZANDO INTELIGÊNCIA ARTIFICIAL PÉROLA (LLM)

Rodrigo Hiroto Morais, Luiza Volfa, Ricardo Evandro Pereira de Oliveira.
Centro de Estudos e Pesquisas Dr. João Amorim (CEJAM), Gestão da Informação, São Paulo, Brasil

Eixo Temático: LT5 - Sistemas Digitais e Dados em Saúde

Introdução

A Linha de Cuidado da Pessoa com Diabetes acompanha cerca de 40 mil pacientes cadastrados no programa no extremo sul de São Paulo, território de alta vulnerabilidade social. Embora o controle metabólico medido pela hemoglobina glicada seja um resultado crucial no acompanhamento da deterioração clínica da doença, ele raramente é lançado no campo estruturado do prontuário e ainda não existe integração com os sistemas dos laboratórios. Por conta dessa falta de resultados, desde 2022 a linha monitora somente as solicitações de exames.

Identificamos que a parte majoritária dos resultados de exames são inseridos em texto livre da evolução clínica sem nenhum tipo de estrutura ou padrão de nomenclatura, o que acarreta na pouca disponibilidade desses dados para análise e acompanhamento dos pacientes. Definimos aqui uma aplicação da arquitetura Pérola para a extração desses resultados de exames com o objetivo de melhorar o atendimento aos pacientes em acompanhamento na linha de cuidado da pessoa com diabetes

Objetivo

Extrair com um grande modelo de linguagem (LLM) os resultados de hemoglobina glicada registrados em texto livre e medir o ganho em volume de dados frente ao campo estruturado.

Métodos

Os dados de evoluções clínicas são replicados na arquitetura medalhão de processamento de dados através do *Azure Data Factory*. A arquitetura reside em lakehouse na Azure, e transformações são aplicadas nos dados com *PySpark no Databricks*.

O modelo GPT-5-Nano é empregado na camada ouro da arquitetura no Databricks com a função *ai_query*. As extrações são feitas apenas nas evoluções clínicas que: possuem mais de 20 caracteres; foram feitas em AME, UBS ou hospital dia; estão associadas a pacientes cadastrados como diabéticos.

A LLM retorna texto em JSON, que é formatado, de-duplicado (duas evoluções podem conter o mesmo exame), e filtrado de valores menores que 3.5 ou maiores que 20 e de resultados com datas maiores que a data da evolução. Nos casos duplicados, damos preferência às evoluções com datas mais recentes.

Desenvolvemos um prompt *few-shot* em colaboração com profissionais da saúde. As evoluções possuem escrita heterogênea, com cada profissional escrevendo os resultados de exames de maneira diferente. Por isso, também foi necessário a identificação de uma série de palavras-chave que se referem ao exame de HbA1c, e também de palavras-chave que se assemelham, mas não configuram uma instância do exame.

hba1c em 04/05/2026: 7,2 hba1c em 19/12/2025 - 7,3
exames(20/1/25): hb glic: 11,7%,
últimos exames laboratoriais em março/25 com resultado de glicada 10,2%
mar/24 cpk 35 hba1c 6,8 cr 1,04 fa 122
exame bioquímico: 03/07/2025 hba1c - 6,1.

Fig 1. Heterogeneidade nas transcrições dos resultados de exames de HbA1c. Cada linha foi retirada de um texto de evolução diferente.

Conclusão

O emprego do menor modelo da família GPT-5, em conformidade com a LGPD e aliado a uma metodologia de alta acurácia, viabiliza a busca ativa precoce de casos descompensados. Ao subsidiar a linha de cuidado com dados oportunos, a ferramenta possibilita intervenções clínicas antes do agravamento, prevenindo desfechos desfavoráveis. Essa inteligência territorial preserva vidas e gera sustentabilidade econômica ao SUS, mitigando internações por condições sensíveis à atenção primária decorrentes do descontrole crônico.

Resultados

O desempenho da LLM foi avaliado em uma amostra de 1.000 registros, utilizando o F-score de n-gramas com espaços, que resultou em uma margem de erro de $\pm 3,10\%$, com nível de confiança de 95%, em relação à população total de 510.329 registros entre janeiro de 2025 e setembro de 2026.

Calculamos a acurácia do modelo através de uma técnica adaptada do F-score de n-gramas. Para cada evolução clínica:

$$F = (1 + \beta^2) \frac{P \cdot R}{P \cdot \beta^2 + R} \begin{cases} P = |r \cap h| \cdot |h|^{-1} \\ R = |r \cap h| \cdot |r|^{-1} \end{cases}$$

Onde h é o conjunto de exames na hipótese (extraídos pela LLM) e r é o conjunto de exames na referência. P representa *precision*, que é a porcentagem de exames na referência que também estão na hipótese, enquanto que R é *recall*, a porcentagem de exames na hipótese também na referência. Definimos $\beta=1$, dando um peso igual a P e a R . Cada n-grama foi definido como o conjunto de data e resultado de exame.

O **F-score médio sobre todas as observações clínicas foi de 0,97** (de um máximo de 1), a mediana foi 1 e o desvio padrão foi de 0,14.

Para a avaliação de volumetria, utilizaremos o ano de 2025 como referência.

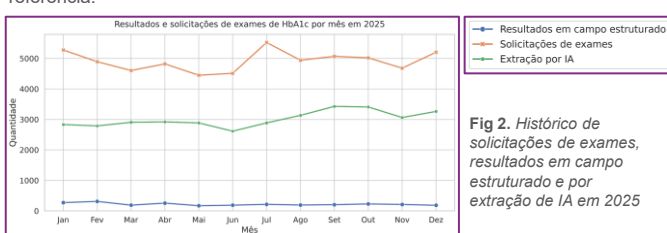


Fig 2. Histórico de solicitações de exames, resultados em campo estruturado e por extração de IA em 2025

Em 2025, dos 30.617 exames de Hemoglobina Glicada solicitados, 2.638 tiveram resultados registrados em campos estruturados (apenas 8,6% dos pedidos). No mesmo período, foram identificadas 310.185 evoluções clínicas que poderiam ter resultados de exames. Com IA, extraímos 36.127 exames (após filtragem e deduplicação), representando um **aumento de 1.369% no volume de dados disponíveis** e 17,9% mais resultados que solicitações.

Constatamos uma média de 0,11 resultados válidos por evolução clínica, concentrados em 27.239 prontuários (média de 1,32 exames por prontuário contemplado), evidenciando que, embora a densidade global seja baixa, o subconjunto de evoluções com dados é denso.

No ano de referência, 19.002 pacientes têm ao menos um resultado extraído, e 20,82% deles têm três ou mais, série que contribui para a análise histórica longitudinal do paciente, que o campo estruturado não oferecia.

Contagem de resultados	Número de evoluções	Percentual das evoluções	Número de pacientes
1	40.068	75,40%	10.568
2	7.181	13,51%	4.477
3 ou mais	5.890	11,08%	3.958

Fig 3. Número e percentual de evoluções clínicas por contagem de resultados extraídos

O custo operacional hoje para cada resultado validado extraído é de R\$ 0.004 em tokens do GPT-5-Nano. A extração sobre o histórico completo das evoluções clínicas de 2025 custou por volta de R\$ 1.500.

Acesse o trabalho na íntegra!

