

O PAPEL DA MATEMÁTICA BÁSICA EM BANCOS DE DADOS VETORIAIS: CASO DO AGRÔNOMO VIRTUAL

Relato de Experiência Monitoria de Disciplina — MD

André Nascimento VELLOSO
Ricardo Ramos SILVA
Fatec Cotia

Resumo: Este trabalho descreve a experiência de monitoria na disciplina de Matemática Básica no curso de Ciência de Dados, com foco na aplicação dos conceitos matemáticos em um projeto prático. O objetivo da monitoria foi auxiliar os alunos na compreensão de temas fundamentais, como operações com números reais, porcentagem e sistemas lineares, aproximando-os da prática no campo da Ciência de Dados. A metodologia consistiu na resolução de exercícios, revisão de conteúdos e contextualização dos conceitos, além da participação no desenvolvimento do sistema Agrônomo Virtual, que utiliza o modelo RAG para recuperação de informações em um banco de dados vetorial Qdrant. Como resultados, teve uma contribuição direta para a escolha e otimização da tecnologia utilizada no projeto. Nas considerações finais, destaca-se o impacto positivo da monitoria na integração dos conceitos teóricos com as necessidades práticas dos trabalhos do curso.

Palavras-chave: matemática básica; metodologia; RAG, monitoria de disciplina.

1. INTRODUÇÃO

A monitoria da disciplina de Matemática Básica no curso de Ciência de Dados tem como objetivo auxiliar os alunos no entendimento dos conceitos fundamentais da área, promovendo uma base sólida para o sucesso acadêmico. Durante minha atuação como monitor, pude perceber que a matemática não se limita a exercícios abstratos, mas possui aplicações diretas em tecnologias utilizadas no campo da ciência de dados.

Essa visão se consolidou quando participei do projeto SMART B100, no qual atuo no desenvolvimento do Agrônomo Virtual, um sistema que utiliza o modelo RAG (Retrieval-Augmented Generation) para recuperar informações técnicas de um banco de dados vetorial Qdrant. O Qdrant armazena dados como vetores e utiliza métricas matemáticas, como a distância euclidiana, para calcular a similaridade entre as perguntas dos usuários e os documentos armazenados.

Neste trabalho, discuto como os conhecimentos de Matemática Básica foram aplicados para avaliar o desempenho do Qdrant, utilizando a distância euclidiana para medir a semelhança entre vetores. A análise desse banco de dados vetorial revela como conceitos matemáticos essenciais são fundamentais para a avaliação e otimização de tecnologias de recuperação de informação.

2. METODOLOGIA

Durante o segundo semestre de 2024, atuei como monitor da disciplina de Matemática Básica no curso de Ciência de Dados, auxiliando os alunos na compreensão de conceitos como operações com números reais, porcentagem e sistemas lineares. A monitoria focou na resolução de exercícios, revisão de conteúdos e contextualização dos conceitos, aproximando-os de aplicações práticas no cotidiano e em outras disciplinas do curso.

Simultaneamente, participei do projeto SMART B100, que tem como objetivo a construção do sistema Agrônomo Virtual, utilizando o modelo RAG (Retrieval-Augmented Generation). Esse modelo "usa a sequência de entrada x para recuperar documentos de texto z e usá-los como contexto adicional ao gerar

a sequência destinada y "¹ (LEWIS et al., 2020) para suportar a busca de informações. Para o armazenamento e gestão dos dados, foi utilizado o Qdrant, um banco de dados vetorial, o banco de dados vetorial em sua definição é especializado na "gestão eficiente de dados vetoriais de elevada dimensão"² (TAIPALUS, 2024), essencial para representar semanticamente textos e outros dados, permitindo a avaliação da similaridade entre eles.

O Qdrant foi escolhido pela sua eficiência em indexar grandes volumes de vetores e seu suporte a filtros de busca semântica, sendo compatível com aplicações baseadas em RAG. Durante o desenvolvimento do projeto, apliquei conhecimentos de Matemática Básica para avaliar o desempenho do Qdrant, especialmente no cálculo da distância euclidiana, uma fórmula matemática fundamental para determinar a semelhança entre vetores em um espaço n -dimensional. A fórmula para calcular a distância euclidiana entre dois pontos x e y é dada por:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Essa medida foi crucial para entender como o Qdrant calcula a semelhança entre os vetores das perguntas dos usuários e os vetores dos documentos armazenados, ajudando a determinar a proximidade entre eles em um espaço de alta dimensionalidade, essencial para o processo de recuperação de informações. (HUANG, 2008)

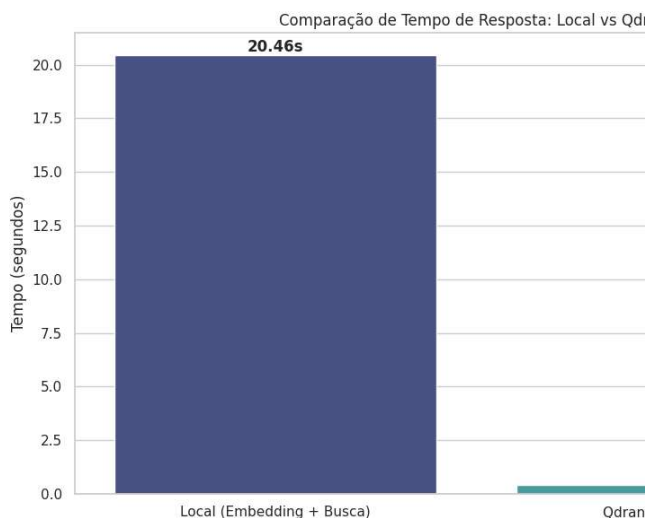
3. RESULTADOS E DISCUSSÕES

Para avaliar como o banco de dados vetorial ajudaria no desenvolvimento do Agrônomo Virtual, realizei uma comparação entre dois cenários distintos: um utilizando código Python para vetorização e busca euclidiana, e outro utilizando o Qdrant, onde os vetores já estão pré-calculados e armazenados.

Gráfico 1: Comparação de Tempo de Resposta entre Vetorização Manual e Banco de Dados Vetorial Qdrant

¹ No original: "RAG models, which use the input sequence x to retrieve text documents z and use them as additional context when generating the target sequence y " (LEWIS et al., 2020)

² No original: "A vector database management system is a specialized type of database management system that focuses primarily on the efficient management of high-dimensional vector data." (TAIPALUS, pg.2, 2024)



Fonte: Autoria Própria

A comparação mostra que a vetorização Local em Python aumenta o tempo de resposta devido à carga computacional, enquanto o Qdrant é mais rápido por já ter os vetores prontos.

4. CONSIDERAÇÕES FINAIS

Em conclusão, minha experiência como monitor de Matemática Básica contribuiu para a aplicação prática dos conceitos matemáticos no desenvolvimento do sistema Agrônomo Virtual. Através da monitoria, foi possível contextualizar a Matemática de

maneira prática para os alunos, ao mesmo tempo em que pude aplicar esses conceitos no uso do Qdrant. O trabalho de monitoria proporcionou uma interação entre teoria e prática, permitindo uma abordagem mais concreta e aplicável aos desafios do dia a dia no campo da Ciência de Dados.

REFERÊNCIAS

HUANG, Anna. Similarity measures for text document clustering. In: NZCSRSC 2008, April 2008, Christchurch, New Zealand. Disponível em: https://www.academia.edu/24048440/Similarity_Measures_for_Text_Document_Clustering. Acesso em: 22 abr. 2025.

LEWIS, Patrick et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv preprint, arXiv:2005.11401, 2020. Disponível em: <https://arxiv.org/abs/2005.11401>. Acesso em: 22 abr. 2025.

Qdrant. Retrieval-Augmented Generation (RAG). Qdrant, 2024. Página da web. Disponível em: <https://qdrant.tech/rag/#>. Acesso em: 22 abr. 2025.

TAIPALUS, Toni. Vector database management systems: Fundamental concepts, use-cases, and current challenges. Cognitive Systems Research, v. 85, p. 101216, 2024. ISSN 1389-0417. Disponível em: <https://doi.org/10.1016/j.cogsys.2024.101216>. Acesso em: 22 abr. 2025.