Score: 15 out of 15 points
Full points awarded. This continues to be an excellent (and ambitious) course project. The report meets all requirements and, along with the GitHub repository, demonstrates meaningful progress toward project goals.

Regarding Path A v Path B: This is already an ambitious project that exceeds course project requirements. There is no need to expand scope.

Strengths:

* Clearly written and well structured report.

* Well-organized GitHub repository demonstrates significant effort and progress toward project goals.

* Report and artifacts demonstrate clear application of course concepts.

Suggested improvements:

* The manually labeled data wasn't available for review. For the final report, consider including the labeled data and more information about the labeling process including how "relevance" is defined and evaluated.

Minor issues:

* GitHub username incorrect in cloning instructions (wjd4) in README.md

* 01_download_pubmed_consciousness.R contains hardcoded path

Additional comments:

* This is more of an FYI since you already have a working acquisition pipeline, but the Entrez CLI is an effective and reliable way to download PubMed data. For example, the following downloads article metadata in PubMed XML format and implements rate limits internally:

esearch -db pubmed -query "consciousness" | efetch -format xml > results.xml

* PubMed has a fairly extensive query language that would allow you, among other things, to query/limit on specific MeSH headings. You are doing this with your classifier, but this might be a quick way to eliminate a large number of articles that are clearly not relevant. See https://pubmed.ncbi.nlm.nih.gov/advanced/. Constructed queries should be usable via Entrez.

Feel free to join OH or reach out via Campuswire for additional feedback or clarifications.