THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

University of Chicago Library

The University of Chicago Library is facing a storage crisis, adding 80,000 volumes annually to its already overflowing collection. To address this, the library engaged the Data Science Clinic to develop a systematic method for identifying materials to relocate to off-site storage.

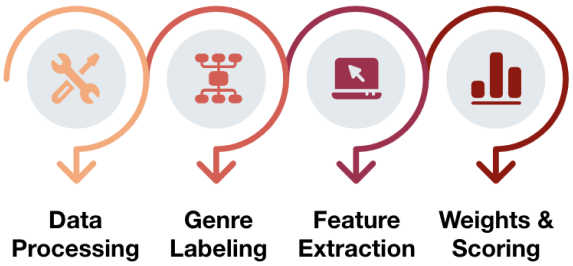

Figure 1: Left-to-right diagram of the analysis and recommendation process.

First, the team processed, organized, and compressed two large datasets: collections data, detailing a large portion of the library's physical and electronic records, and usage data, tracking checkouts over the past decade.

Next, the team utilized pre-classified genre information for 97% of the records to support high-level analysis. Grouping books by genre, alongside other features, enabled the development of a weighted scoring algorithm that ranks volumes by importance for on-site storage. The team then began evaluating an initial list of one million books recommended for relocation. Some analysis on their genres can be seen below.

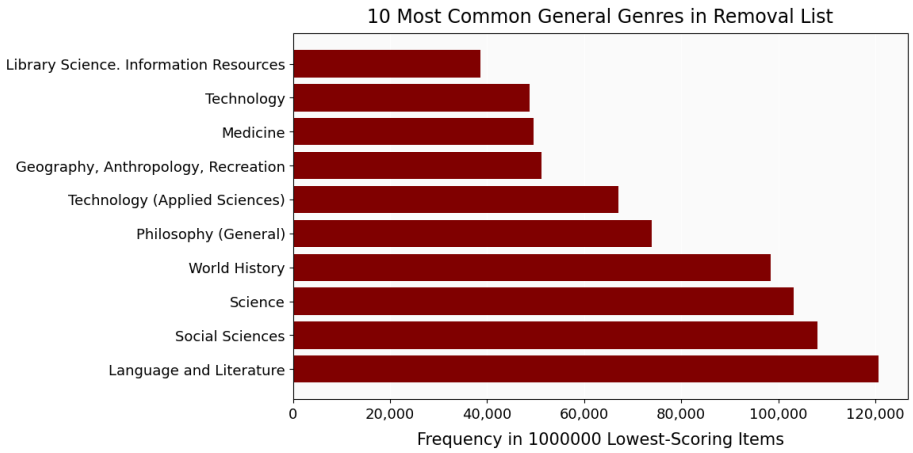

Figure 2: Top 10 genres of items initially recommended for off-site storage, with genre selection based on a combination of low usage, electronic availability, and collection size. 12% of books on the removal list are in "Language and Literature".

In partnership with the library, the team plans to optimize the scoring algorithm over the next quarter. Overall, the team has established a replicable process that provides the foundation for a reusable tool to assist with the library's storage issues in the future.