

Alert Collector

Matthew Galloway, Col. Ed.

Using Statistics and Machine Learning to Manage Foreign Language Collections, Programs, and Services in Public Libraries

Yelena G. Nedlina

The future of foreign language departments is likely to involve moving beyond basic spreadsheets and toward advanced data platforms (such as Tableau or Power BI) that deliver real-time, actionable insights. By combining traditional usage statistics with machine learning, libraries can build more inclusive, responsive, and truly global collections and services.

Managing Foreign Language Collections with Machine Learning: Analysis and Predicting Future Needs

Statistics already provide valuable guidance for decision-making, especially in collection development. One example is evidence-based selection planning, which uses sophisticated machine learning and behavioral data to help librarians make more informed selection decisions. Common types of data-driven collection development include turnover rate analysis, gap analysis, and usage pattern analysis.

Turnover Rate Analysis

By calculating turnover rate (total circulation divided by collection size), librarians can identify high-demand languages and genres that may require increased investment. For example, periodicals, general adult fiction, biographies, detective fiction, and romance are traditionally popular among older adult patrons; these genres often need ongoing purchasing to maintain an adequate supply.

Gap Analysis

Comparing circulation data with demographic data can reveal collection gaps (i.e., a large population of speakers of a particular language in the area but disproportionately low circulation of materials in that language). In such cases, the next step is to investigate why the gap exists. Possible causes might include limited public transportation to the library, a collection that is too small or

not appealing to the community, the absence of a staff specialist to develop culturally relevant programs, a lack of outreach and public relations, or other barriers. The next step is to address the issue and then collect and analyze new data to determine whether circulation improves.

Usage Pattern Analysis

By analyzing user behavior alongside demographic data, libraries can personalize, optimize, and diversify foreign language collections to better serve multicultural communities. For example, Russian alphabet books and fairy tales are especially popular among Russian-speaking families with children, particularly during winter, when seasonal holidays and family gatherings make reading and celebrations more common.¹ Recognizing these patterns helps librarians design timely displays, exhibits, programs, and more.

Predictive Analytics: A Powerful Tool for Planning Ahead

Machine learning shifts the focus from describing the past to forecasting future needs. Predictive analytics can support more targeted planning for collections, programs, and services based on the interests and behaviors of local communities as noted by Meesad and Mingkhwan:

Similar to commercial streaming platforms, library AI tools can recommend foreign-language items based on a patron's borrowing history, helping users discover additional materials in their native language. By analyzing trends in interaction data from library systems, machine learning models can also suggest which types of foreign literature are likely to be most in demand in upcoming seasons.²

Will Foreign Language Departments in Public Libraries Continue to Exist?

Real-time translation, multilingual chatbots, automated collection curation for diverse languages, and personalized language-learning tools can reduce barriers and improve access for non-native English speakers. For example, multilingual chatbots such as the "T-Rex" system at the University of Calgary use large language models (LLMs) and retrieval-augmented generation (RAG) to provide 24/7 reference support in multiple languages.

With expanding AI tools, it is reasonable to ask: Will public libraries still need separate foreign language departments?

To explore this question, we contacted Milos Markovic, Manager of the International Languages Department, and Victoria Kabo, Subject Librarian (Russian and Slavic languages) at the Cleveland Public Library in Cleveland, Ohio.

Cleveland is a multicultural city, and many libraries in the region maintain foreign language collections and services. The Cleveland Public Library operates a dedicated International Languages Department with the largest and most widely used collection of its kind in the area. Patrons from across Cleveland and beyond visit the department throughout the year, and smaller library systems frequently request materials through interlibrary loan using the CLEVNET catalog.

The International Languages Collection at Cleveland Public Library reflects the city's diverse global community. We are beginning a digitization program to make unique and valuable items from our research collections available online for our readers.³

The department is known not only for its extensive collection but also for its strong cultural programming, which has made it highly visible and well recognized in the community. The department offers numerous cultural events, including holiday celebrations, author talks, book discussions, and other interest-based gatherings for speakers of more than 50 languages.

We provide, and plan to continue to provide, cultural programs to our patrons. Our library (and our department as part of the library) fully shares and follows the recommendations of the American Library Association: "The American Library Association (ALA) supports public libraries that create inclusive and relevant cultural programs. These programs can help build community, promote understanding, and celebrate diversity."⁴

Future Directions for Data Collection and Machine Learning in Public Libraries

Given the growing role of statistics and machine learning in many fields—including librarianship—alongside the continuing need for specialized collections and services for speakers of languages other than English, what are the most promising directions for machine learning use in public library foreign language services?

To address this question, we interviewed Danielle Prilepskiy, a Statistics and Machine Learning student at Carnegie Mellon University.

Prilepskiy stated, "Machine learning is a subset of artificial intelligence focused on building algorithms to make educated predictions. Statistics is centered on modeling data to understand relationships and patterns, foundations which machine learning builds upon. Modern data science lies at the intersection of these fields, using statistical methods to identify the questions machine learning can answer." Since public libraries generate substantial amounts of data, foreign-language departments can definitely benefit from the applications of machine learning. Success would depend on maintaining an organized database, using automation to reduce workload in applicable cases, and involving specialists who can clean the data. Libraries such as Scikit-learn are used for predictive modeling and offer tools like the cosine similarity function, which is particularly useful in batch processing. For example, a public library could apply weighted cosine similarity to identify which books to add to its catalog by analyzing the factors most strongly associated with a book's success. The same approach could also support more personalized and targeted recommendations for readers. With well-implemented models, libraries can develop tools that analyze their data, uncover trends, and guide practical improvements to their systems.⁵

Conclusion

Foreign language departments in public libraries provide essential collections, cultural programming, and services to multilingual communities. By combining traditional usage statistics with machine learning, libraries can strengthen decision-making, improve responsiveness, and better serve patrons who speak languages other than English.

References

1. Yelena Nedlina, "Public Services to Russian-Speaking Patrons: The View of a Bilingual Librarian," *Bookmobile and Outreach Services* 10, no. 1 (2007): 25–31.
2. Phayung Meesad and Anirach Mingkhwan, "Data-Driven Library Management: From Data to Insights," in *Libraries in Transformation* (Springer, 2024), 169–209.

3. "International Languages," Cleveland Public Library, Accessed February 11, 2026, <https://cpl.org/internationallanguages/>.
4. Milos Markovic and Victoria Kabo, personal interview, February 8, 2026.
5. Danielle Prilepskiy, personal interview, February 12, 2026.