ARIA
APPLIED RESEARCH IN ACTION

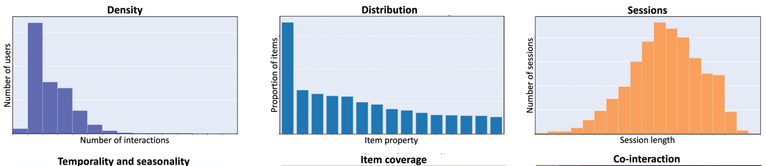
DataStats: An Automated Framework Supporting the Design and Development of Recommender Systems

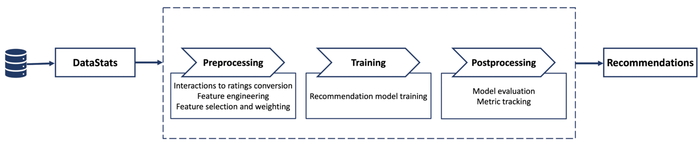
An innovative framework empowering recommender systems development with automated analysis, improved decision-making support, and domain-focused insights for enhanced personalization and efficiency.

Anusha Prabhudev

Eldan Cohen
ACADEMIC SUPERVISOR

Garrin McGoldrick and Marie Al Ghossein
INDUSTRY SUPERVISORS





PROJECT SUMMARY

Developing personalized recommender systems is a complex and challenging task, demanding profound comprehension of the underlying science and the specific data and domain where recommendations are deployed. To address this challenge, we present DataStats, an automated framework designed to streamline analysis, enhance decision-making, and improve the efficiency of recommender system development.

We combine fundamental concepts from recommender systems with practical machine learning expertise. The culmination of this research effort is a novel framework offering a suite of powerful tools to automate the standardized analysis of recommendation datasets. DataStats encompasses key features such as detailed data profiling, data distributions, user journey analysis, feature selection, and session-based evaluation, contributing to its multifaceted efficacy. It incorporates interactive visualizations, comprehensive metrics, and descriptive summaries to empower decision-making with actionable insights, ensure data quality, and guide pipeline design, consequently improving client onboarding efficiency and elevating decision quality.

Employing user-centric design, advanced data analysis, and machine learning techniques, DataStats is successfully implemented as automated notebooks, enabling seamless integration into recommender system pipelines. The framework's deployment demonstrates its effectiveness in optimizing the client onboarding process, notably reducing the onboarding time by 20%. End-user evaluation and feedback further testify to the practical utility of DataStats in enhancing the recommender system development process.

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing