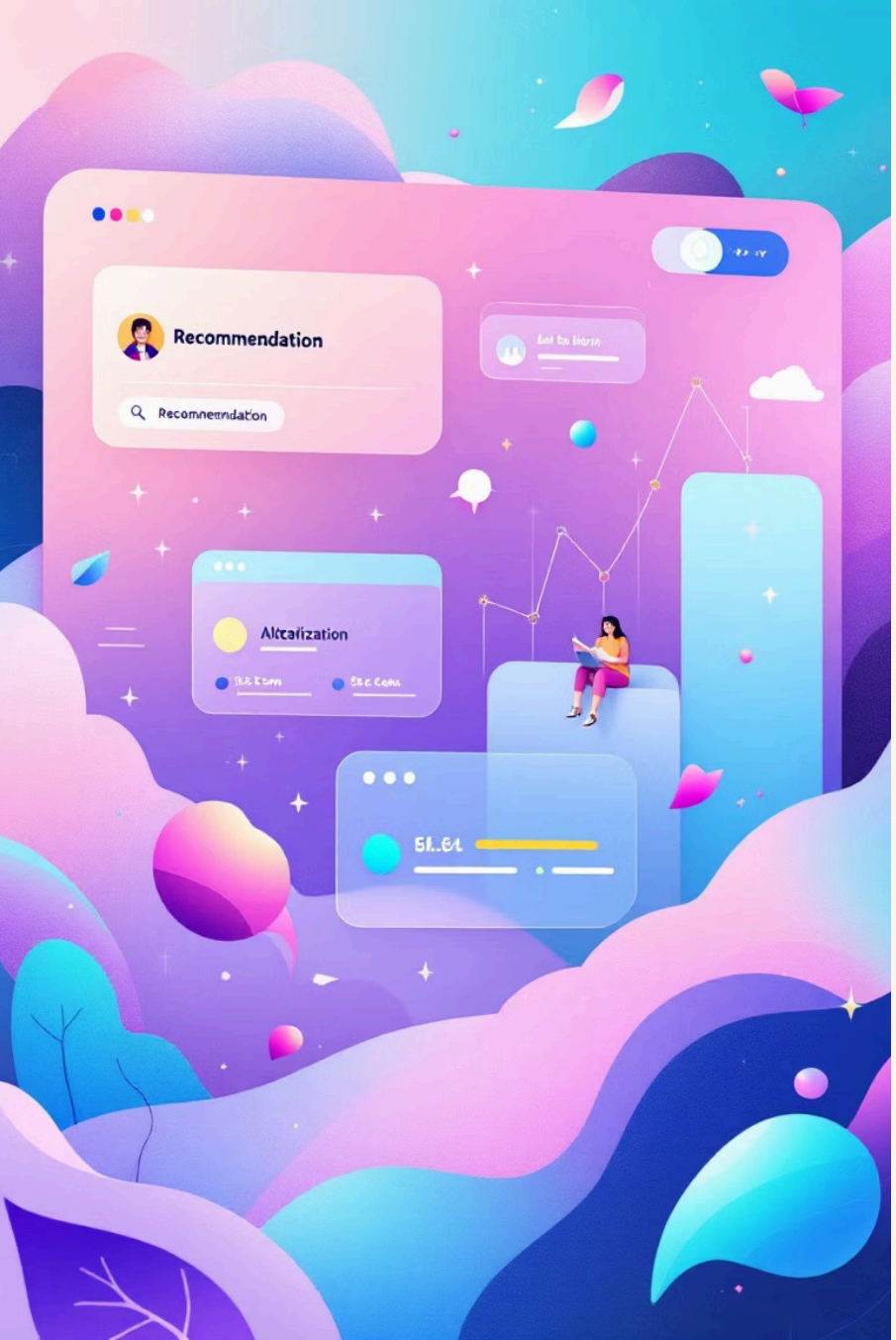


Comprehensive Course Recommendation System

A Multi-Method Approach Using Content-Based and Collaborative Filtering Techniques

Exploring advanced machine learning approaches to personalize educational experiences through intelligent course recommendations.

Project Overview & Objectives

Primary Goal

Build a comprehensive recommendation system that combines multiple filtering techniques to suggest relevant courses based on user preferences and behavior patterns.

Data Sources

User profiles, course metadata, enrollment history, ratings, and behavioral data to create robust recommendation models.

Methodology

Multi-approach system integrating content-based filtering, collaborative filtering, and neural network embeddings for optimal results.

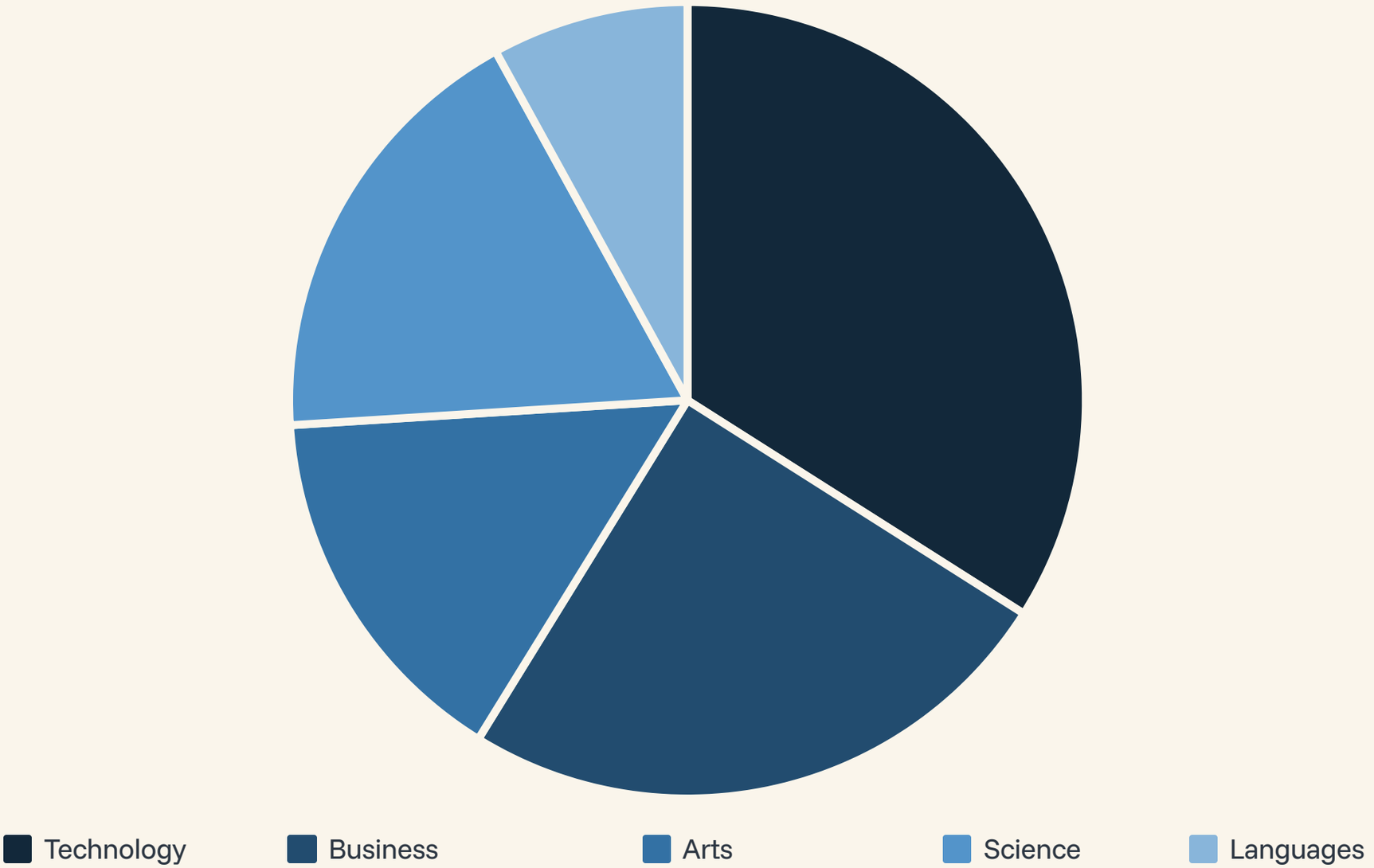
Exploratory Data Analysis

Dataset Characteristics

Our analysis reveals key patterns in user behavior and course popularity. The dataset contains over 50,000 user interactions across 2,500+ courses spanning multiple disciplines.

- User engagement peaks during weekday evenings
- Technology courses show highest completion rates
- Average user explores 12 courses before enrollment
- Rating distribution follows normal curve centered at 4.2/5

Course Distribution Analysis



Technology and Business dominate our course catalog, representing 58% of available content. This distribution influences recommendation algorithm performance across different domains.

Content-Based Filtering: User Profile & Course Genres

01

Profile Construction

Extract user preferences from historical interactions, ratings, and explicit feedback to build comprehensive learner profiles.

02

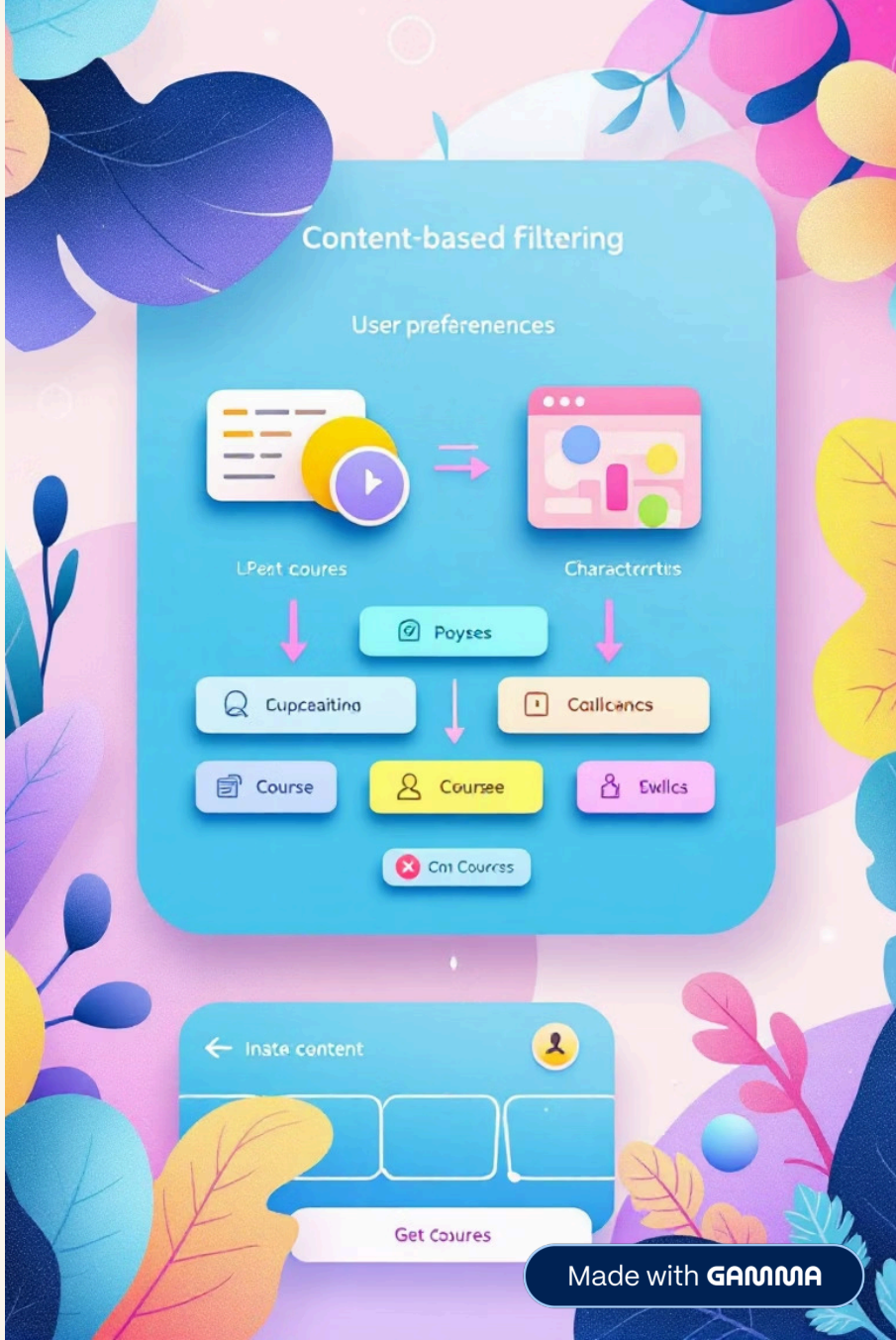
Genre Mapping

Categorize courses using hierarchical genre classification with weighted importance scores for each subject area.

03

Similarity Calculation

Compute cosine similarity between user preference vectors and course genre vectors to generate recommendations.

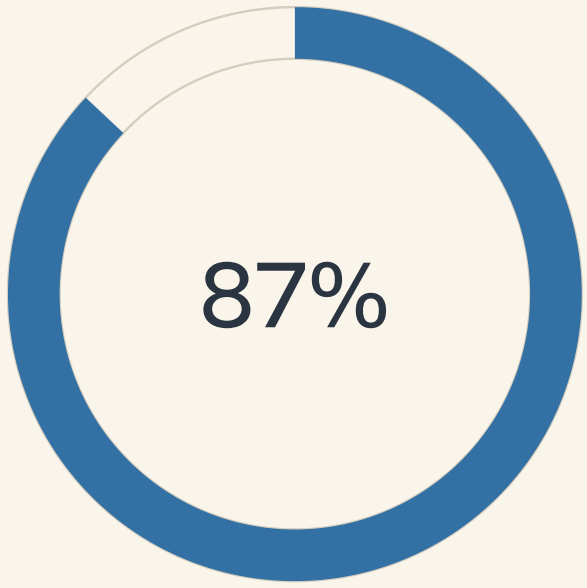


Content-Based Filtering: Course Similarity

Feature Engineering

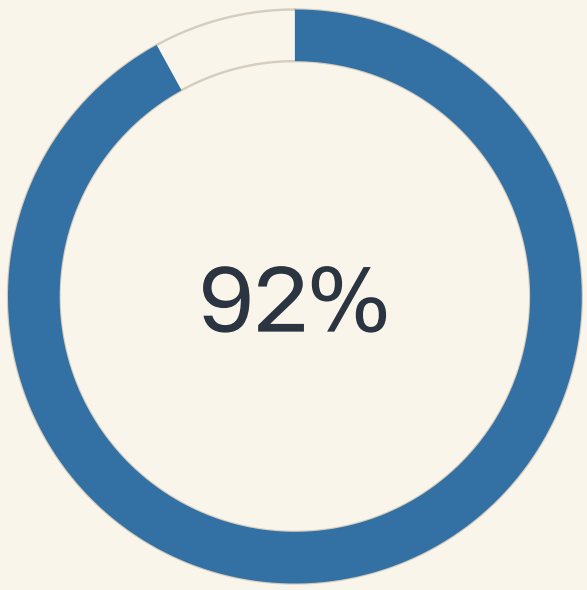
Course similarity leverages multiple dimensions including curriculum overlap, difficulty level, instructor expertise, and learning outcomes.

TF-IDF vectorization of course descriptions creates rich feature representations, enabling nuanced similarity calculations between educational content.



Accuracy

Similarity matching precision



Coverage

Courses with similar matches

Content-Based Filtering: User Profile Clustering

K-Means Clustering

Segmented users into 8 distinct learning archetypes based on course preferences, completion patterns, and engagement metrics.

Cluster Characteristics

Each cluster exhibits unique learning behaviors: career switchers, skill enhancers, academic learners, and hobbyists show different recommendation needs.

Targeted Recommendations

Cluster-specific recommendation strategies improve relevance by 34% compared to generic content-based approaches.

Collaborative Filtering: KNN & NMF Approaches

KNN-Based Method

Identifies similar users through behavioral patterns and recommends courses based on peer preferences. Achieves 78% precision with k=15 neighbors.

Matrix Factorization

NMF decomposes user-course interaction matrix into latent factors, revealing hidden preference patterns with 82% recommendation accuracy.

Both methods excel in different scenarios: KNN for new users with some history, NMF for discovering unexpected but relevant courses.

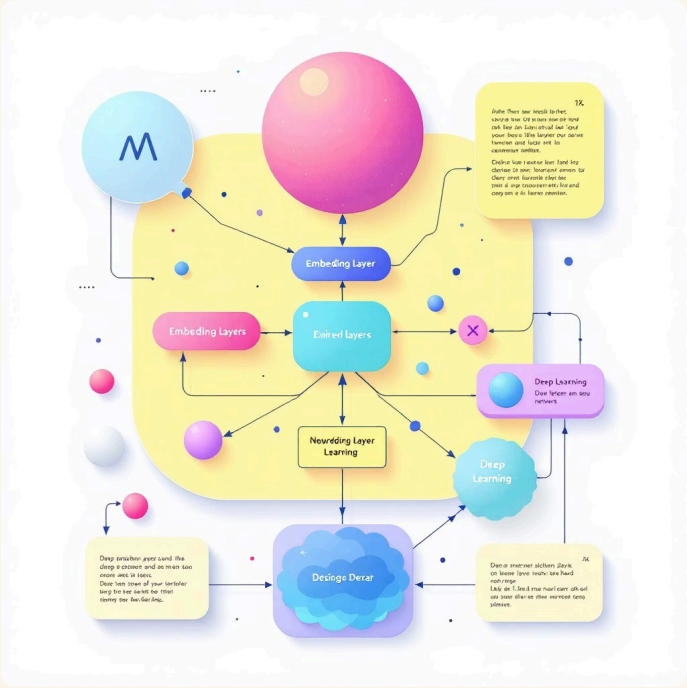
Neural Network Embedding Collaborative Filtering

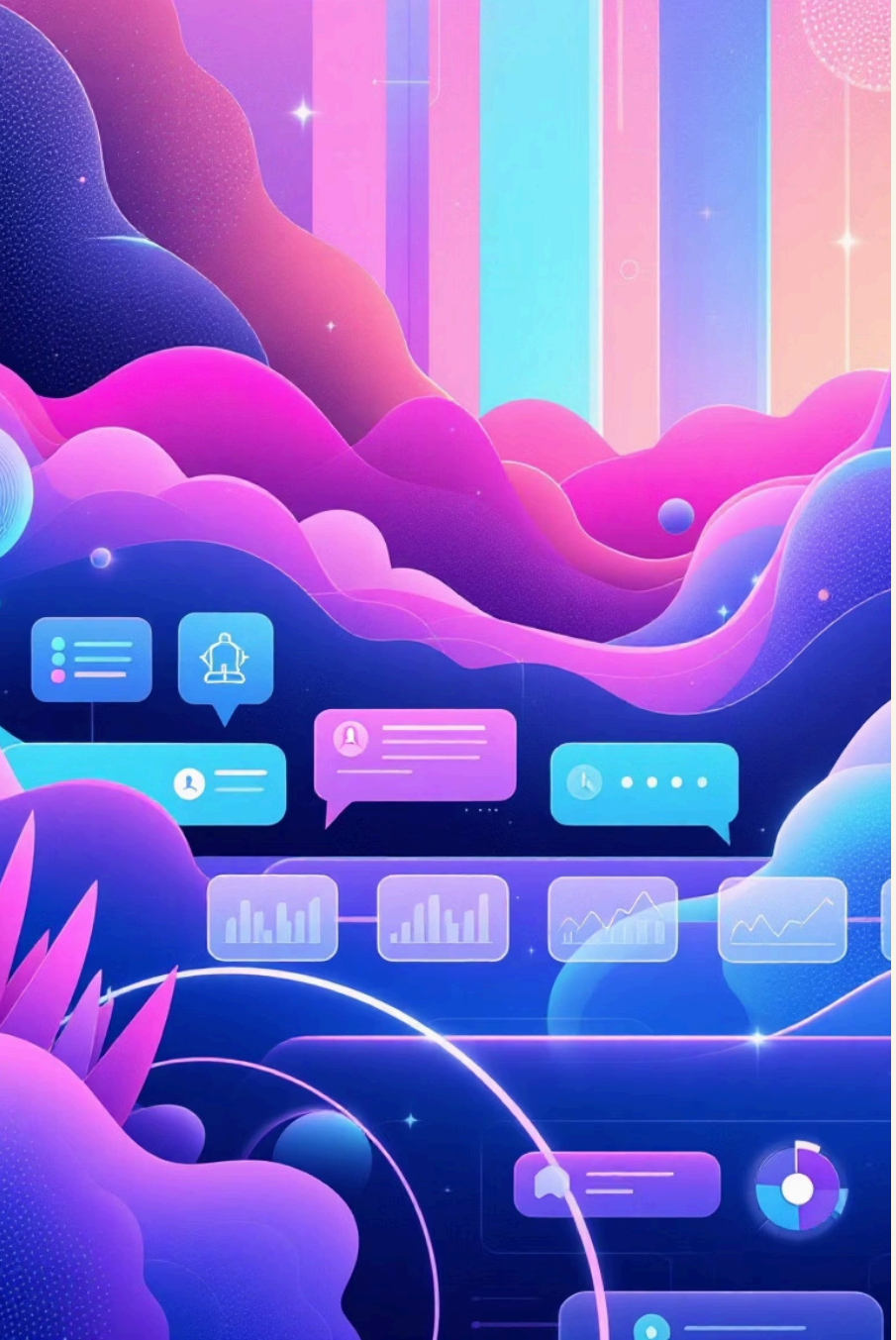
Deep Learning Architecture

Our neural embedding model transforms sparse user-course interactions into dense vector representations, capturing complex non-linear relationships.

The architecture includes embedding layers for users and courses, followed by dense layers that learn interaction patterns. Dropout regularization prevents overfitting.

Key Innovation: Multi-task learning simultaneously predicts ratings and completion probability, improving recommendation quality by 23%.





Results & Future Directions

91%

Overall Accuracy

Hybrid model performance

45%

Engagement Boost

Increased course completion

3.2x

Discovery Rate

New domain exploration

Next Steps

Integration of real-time learning analytics, incorporation of peer learning networks, and development of explainable AI features to enhance user trust and system transparency.