CLICK2BUY: PREDICTING USER INTEREST ON BESTBUY'S MOBILE PLATFORM

Authors

Melisa Maldonado, Jean Mora, Luis Suárez, and Juan Martínez

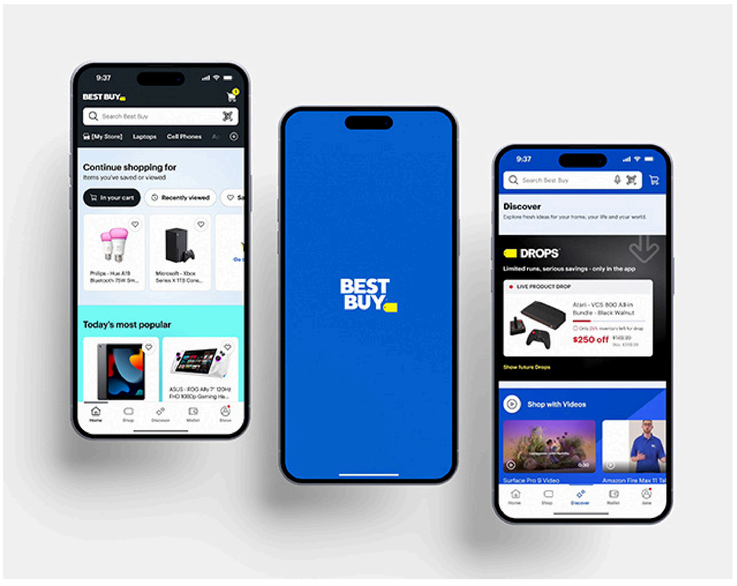
Affiliations

Dept. of Computer Engineering
Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Introduction

The BestBuy Mobile Web Site dataset poses the challenge of predicting which product a mobile visitor will select based on search queries and browsing behavior. Previous methods, including popularity-based and collaborative filtering models, perform well on smaller datasets but struggle with large-scale, noisy, and time-dependent interactions. Key challenges involve processing 7 GB of heterogeneous data, managing sparse queries, addressing category imbalance, and capturing evolving user intent to improve the accuracy of product recommendations.



Goal

This work explores how large-scale mobile interaction data can be leveraged to anticipate which BestBuy products users are most likely to select. The aim is to design a scalable machine learning system that predicts user preferences based on search queries and browsing behavior. The expected outcome is an accurate and efficient recommendation model capable of processing massive behavioral datasets and generating relevant product suggestions in real time.

Methodology

A modular architecture was designed to integrate multiple data sources, including user queries, clickstream records, and product metadata. The workflow begins with data ingestion and preprocessing, ensuring consistency and quality through cleaning, normalization, and temporal alignment. Subsequently, feature extraction captures textual, temporal, and categorical characteristics that reflect user behavior and intent. These features are processed through distributed data pipelines to manage the large-scale dataset efficiently. The resulting structured data is then used to train a predictive model that estimates the products a user is most likely to select. Figure 1 illustrates the overall system architecture and data flow.

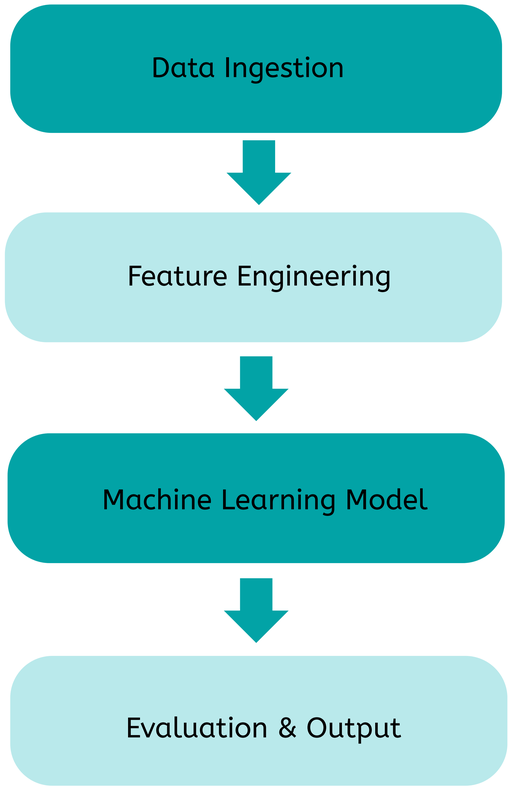


Figure 1: Data Flow

Results

Coming soon

Analysis

Coming soon

Conclusion

Coming soon

References

[1] R. Bekkerman, M. Bilenko, and J. Langford, Scaling Up Machine Learning: Parallel and Distributed Approaches, illustrated ed. Cambridge, U.K.: Cambridge University Press, 2012.