Click2Buy: Predicting User Interest in Best Buy Mobile Website

Juan Diego Martinez Beltran 20231020131
Luis Felipe Suarez Sanchez 20231020033
Melisa Maldonado Melenge 20231020110
Jean Pierre Mora Cepeda 20231020105

25 October, 2025

Business Model

Modern e-commerce platforms like Best Buy handle millions of online interactions daily, each reflecting a user's intent, curiosity, or need. However, predicting what a user truly wants remains a complex challenge.

The ACM SF Chapter Hackathon on Big Data presents a real-world problem: to predict which product a mobile visitor will be most interested in based on their search queries and browsing behavior.

The reason for this challenge lies in the growing complexity of user behavior.
Customers perform short, noisy, and ambiguous searches, and their interests evolve dynamically throughout a session. Traditional recommendation approaches—based solely on popularity or simple keyword matching—struggle to interpret this uncertainty.

Goal

Predict which BestBuy product a mobile web visitor will be most interested in based on their search query or behavior

In essence, this system emerges from the need to transform chaotic, unstructured user data into meaningful insights—creating adaptive, self-regulating models that reflect the complexity of human decision-making online.

Challenges

- **Data Quality and Ambiguity**

User search queries are often incomplete or inconsistent, and the dataset doesn't guarantee that clicks directly result from each query. This uncertainty adds noise and makes intent prediction harder.

- **Time differences between actions**

Users made searches and clicks at different times, and we only knew that the click happened within five minutes after the search.
It was difficult to tell if the click was really related to that search or something else.
We had to be careful when organizing and matching these events.

Input / Output

user	sku	category	query	click_time	query_time
000000df17cd56a5df4a94074e133c9d4739fae3	2125233	abcat0101001	Televisiones Panasonic 50 pulgadas	2011-09-01 23:44:52.533	2011-09-01 23:43:59.752
000001928162247ffaf63185cd8b2a244c78e7c6	2009324	abcat0101001	Sharp	2011-09-05 12:25:37.42	2011-09-05 12:25:01.187
000017f79c2b5da56721f22f9fdd726b13daf8e8	1517163	pcmcat193100050014	nook	2011-08-24 12:56:58.91	2011-08-24 12:55:13.012
000017f79c2b5da56721f22f9fdd726b13daf8e8	2877125	abcat0101001	rca	2011-10-25 07:18:14.722	2011-10-25 07:16:51.759
000017f79c2b5da56721f22f9fdd726b13daf8e8	2877134	abcat0101005	rca	2011-10-25 07:19:51.697	2011-10-25 07:16:51.759

Input

User interaction

- Search Queries
- Reviews
- Purchase History
- Click_time
- Session MetaData

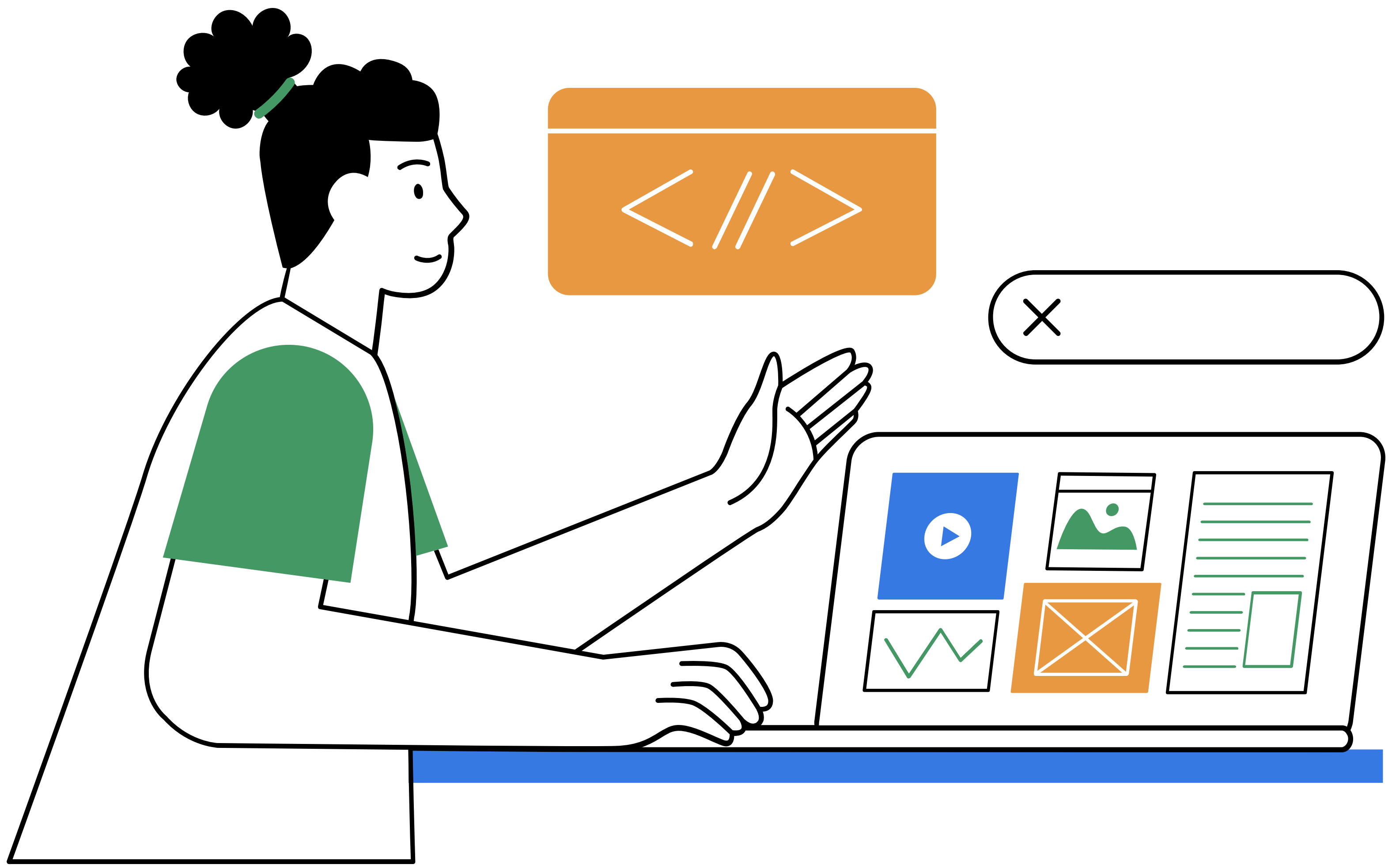
Product Information

- Product Categories

train.csv

Output

- Predicted Product SKUs (Main Prediction)
- Featured products
- Metrics
- User behavior insights



Variation in search queries

A tiny difference in a user's search query or browsing pattern can completely change what product they click.

"iPhone" vs "iPhone case" leads to a totally different set of behaviors.

Multiple account behavior

If the same user has two different accounts and searches for different things in each one, the system may detect them as two separate behavior profiles.

Sensitivity and chaos

Random user intent (gifts, curiosity, price comparison).

Non-linear interactions.

Over two years, user interests, trends, and product availability change constantly.

Open System

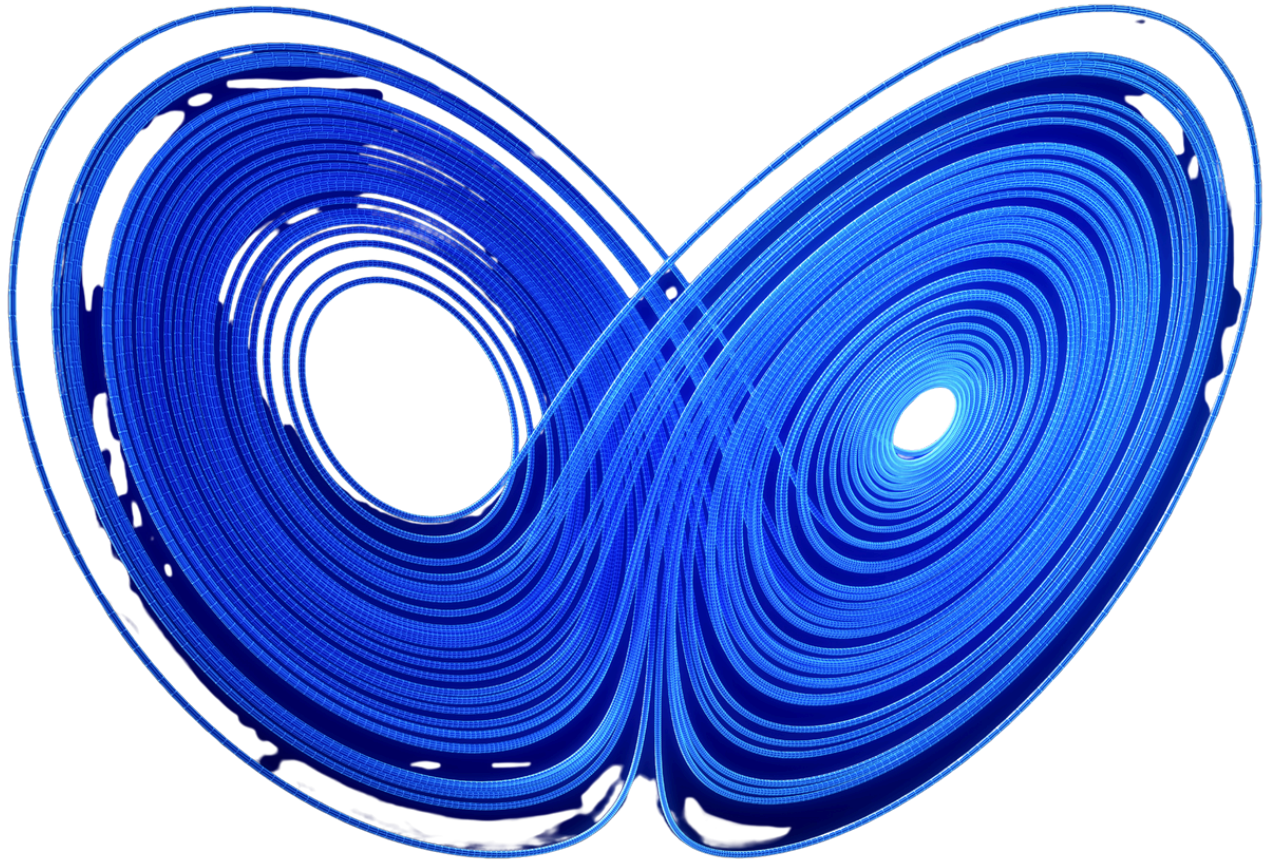
Interacts with its environment: marketing campaigns, competitor actions, new product launches, seasons, and even user emotions.

Data doesn't exist in isolation → External factors that can't be fully controlled or predicted.

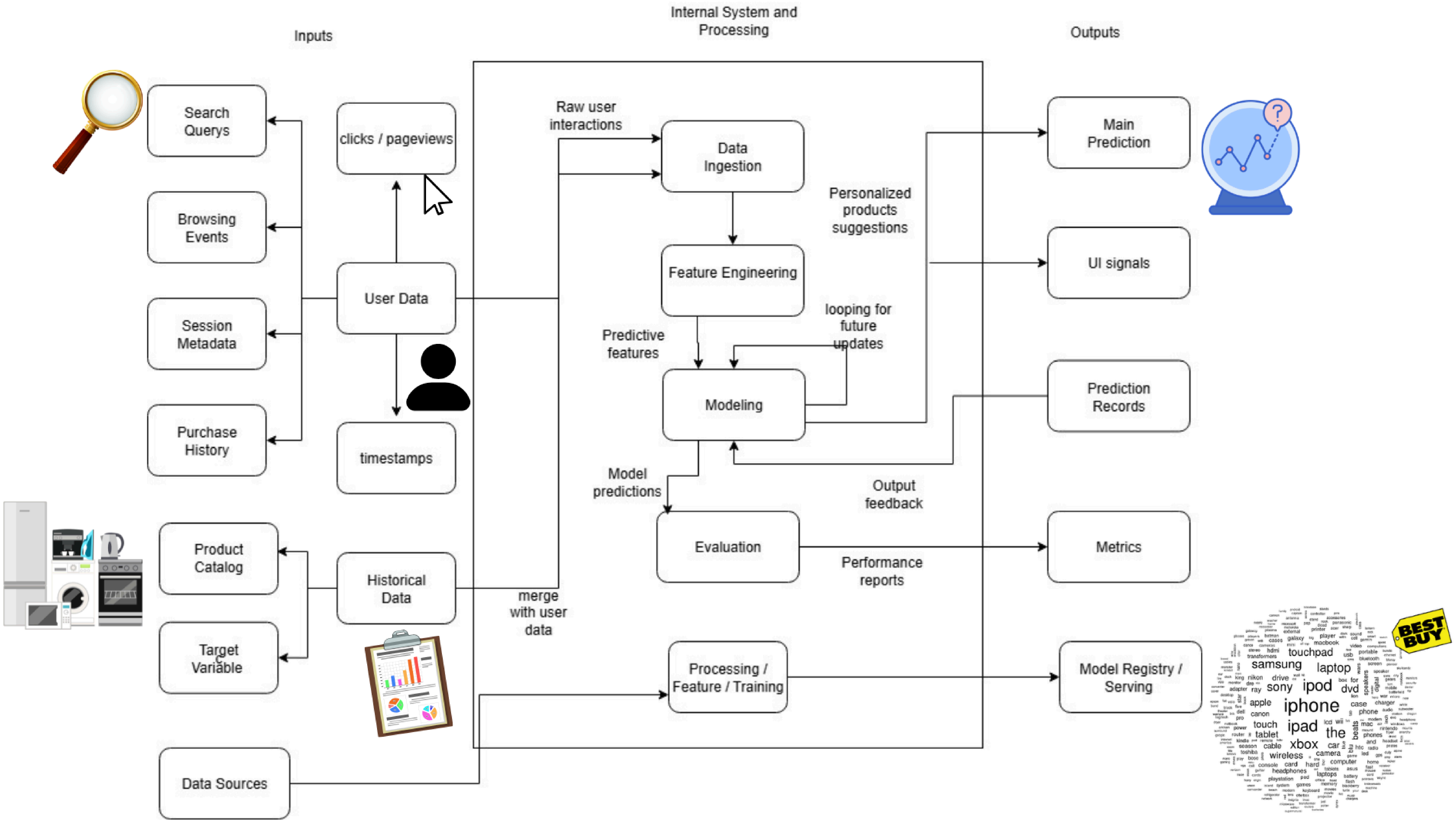
- Visit other websites like Walmart, Amazon, or eBay,
- Users might see different prices or promotions on other websites, so they return to Best Buy to check the prices again. This changes the behavior pattern that the system has been studying for that user.

Chaotic attractor

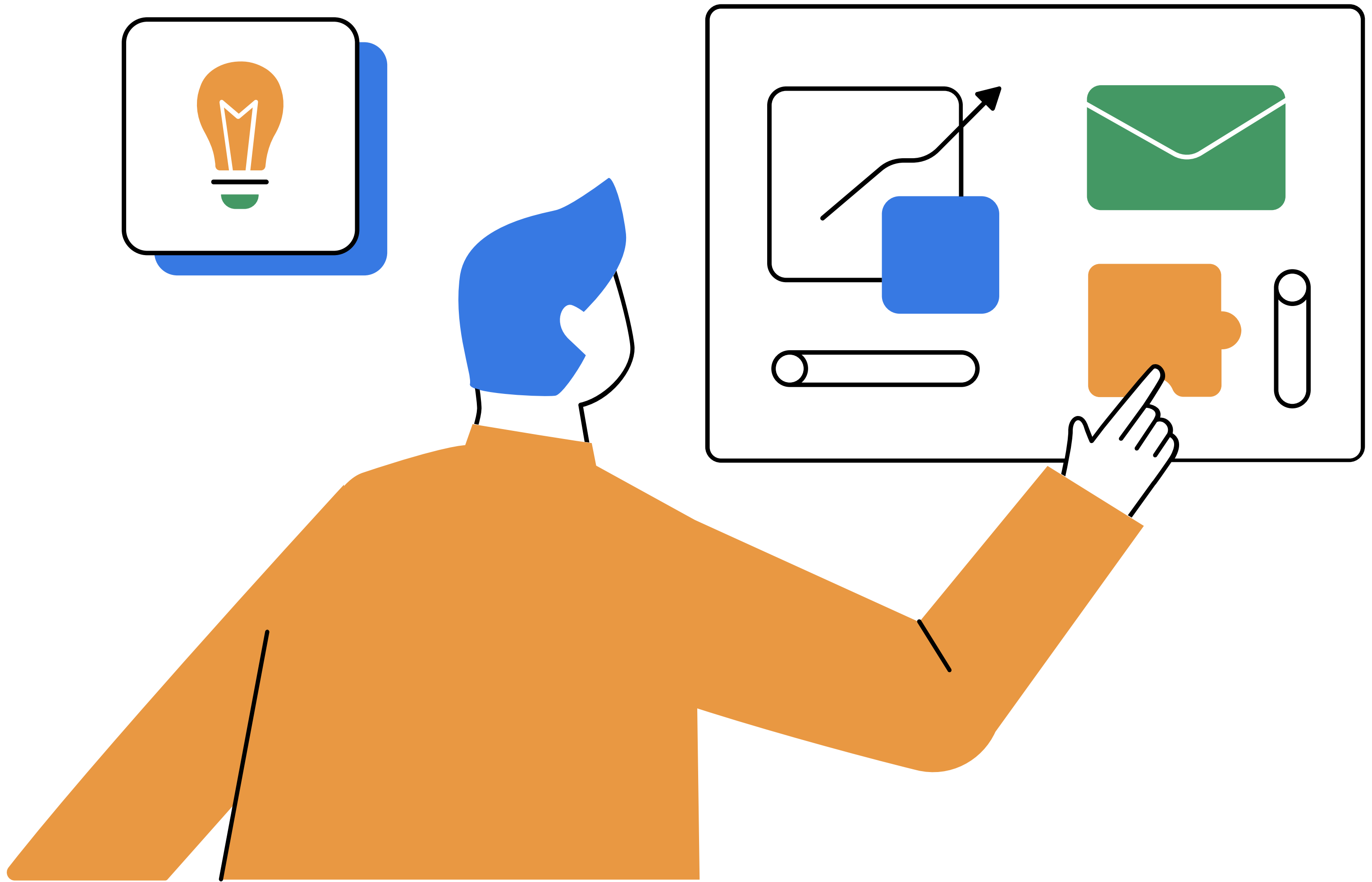
Despite all this chaos, we can start exploring predictions through trends, the first step toward finding the chaotic attractor.



Architecture diagram



Functional requirements



Analysis and Behavioral Understanding

- The system must analyze user behavior to identify browsing trends and preferences across product categories.
- Must process product metadata, including category, brand, and review information.

Evaluation and Feedback

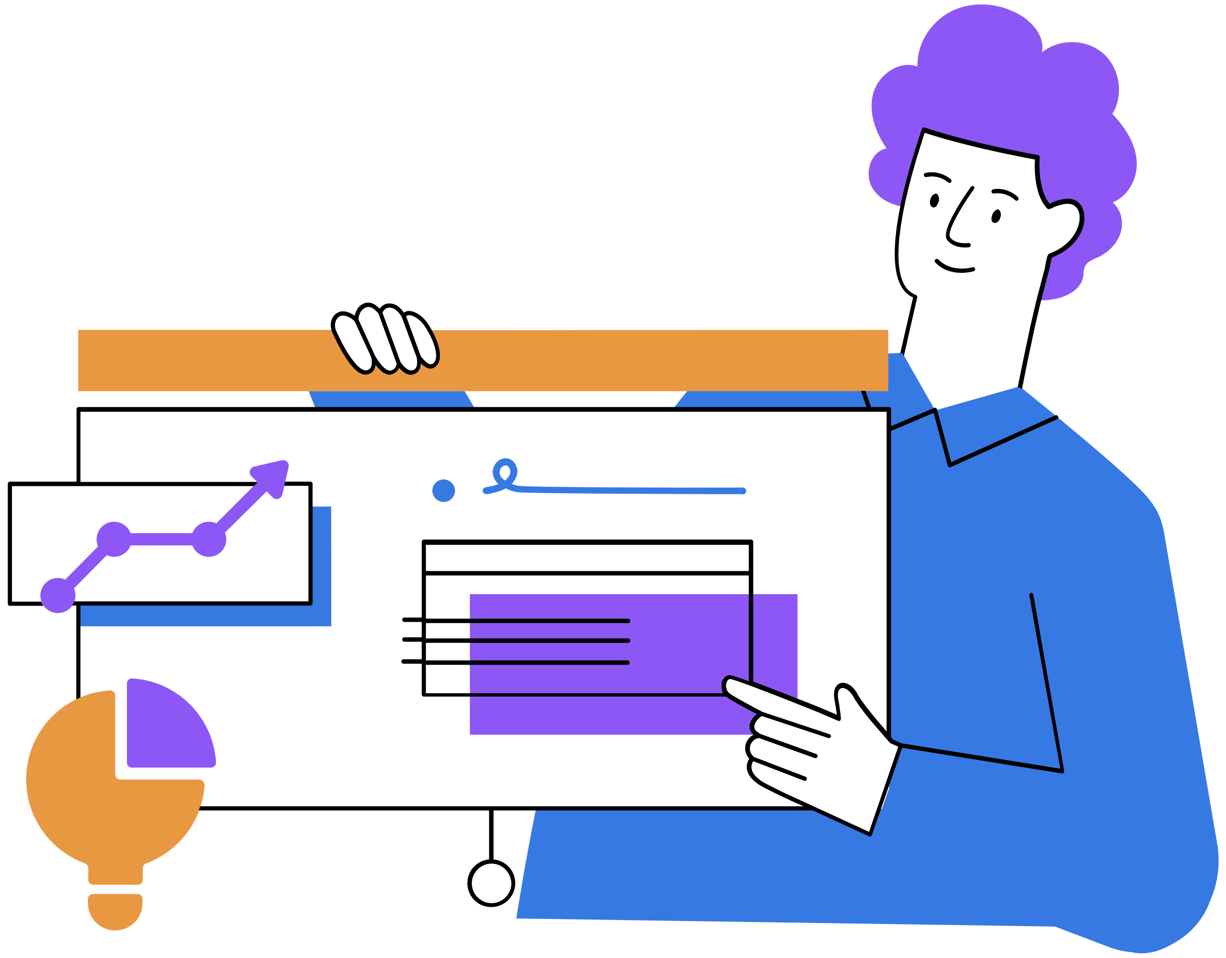
- The system must evaluate model performance using MAP@5 as the main metric. (Sensitivity)

Data Processing & Storage

Must parse and merge multiple data sources:

- train.csv and test.csv for user interactions.
- product_data.tar.gz for product metadata and attributes.

Non-functional requirements



Robustness

- Maintain consistent recommendations even when user behavior is noisy, incomplete, or unusual..

Performance

- Response time: Predictions should be generated efficiently for 1.2M user queries.

Data Size and Storage Limitations

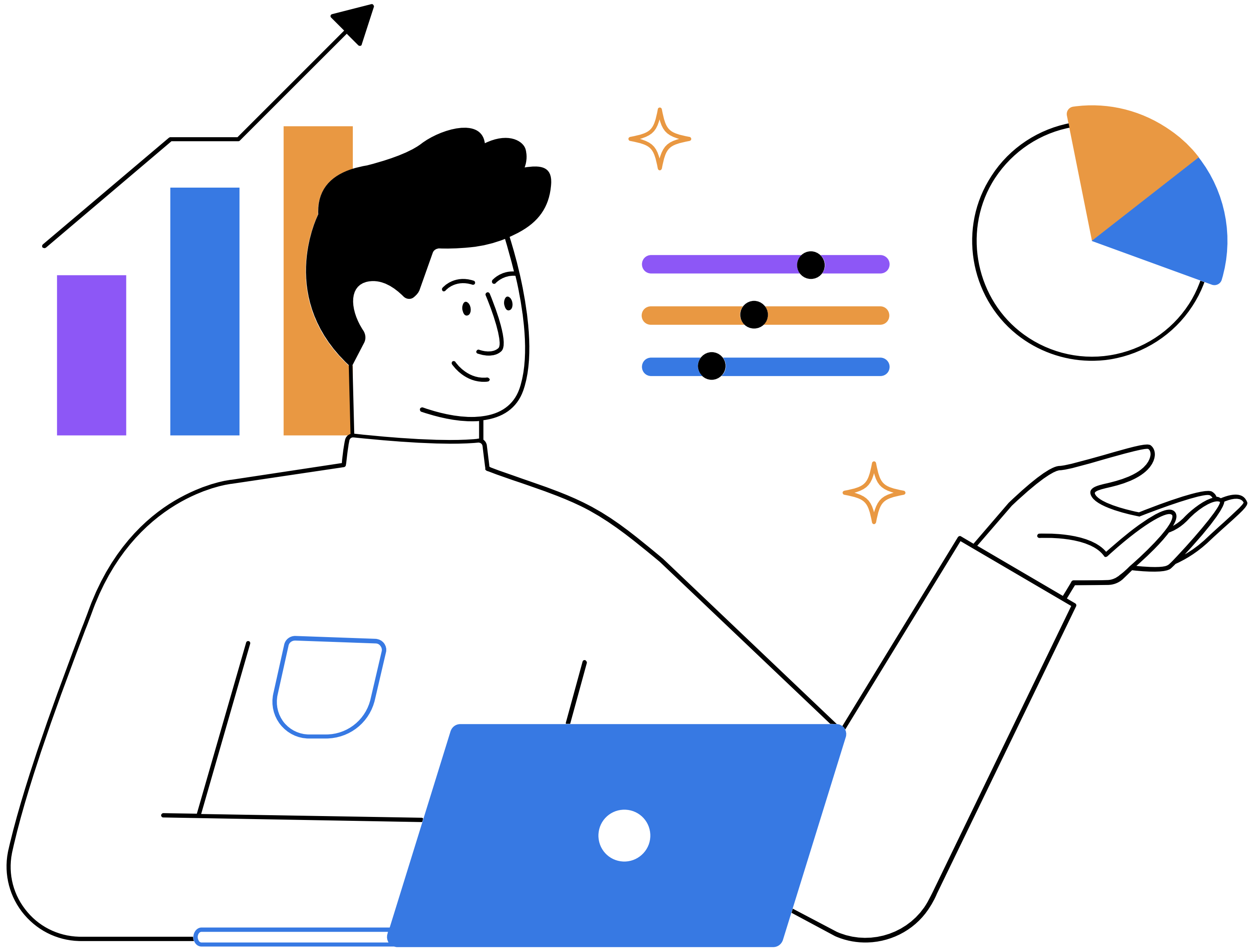
- The dataset is 7 GB, which is large for local machines.
- This restricts how much data can be loaded into memory at once.

Constraints

Limited Temporal Range

- The system can only use information within five minutes between query_time and click_time.
- Longer-term user behavior (hours, days later) is not available.

The project established a modular and scalable framework to predict user interest on Best Buy's mobile platform using behavioral and textual data. However, due to computational and resource limitations, the full deployment and empirical validation of the model remain pending. The design phase revealed that managing 7 GB of data requires distributed computation and careful preprocessing to ensure consistent and interpretable results.

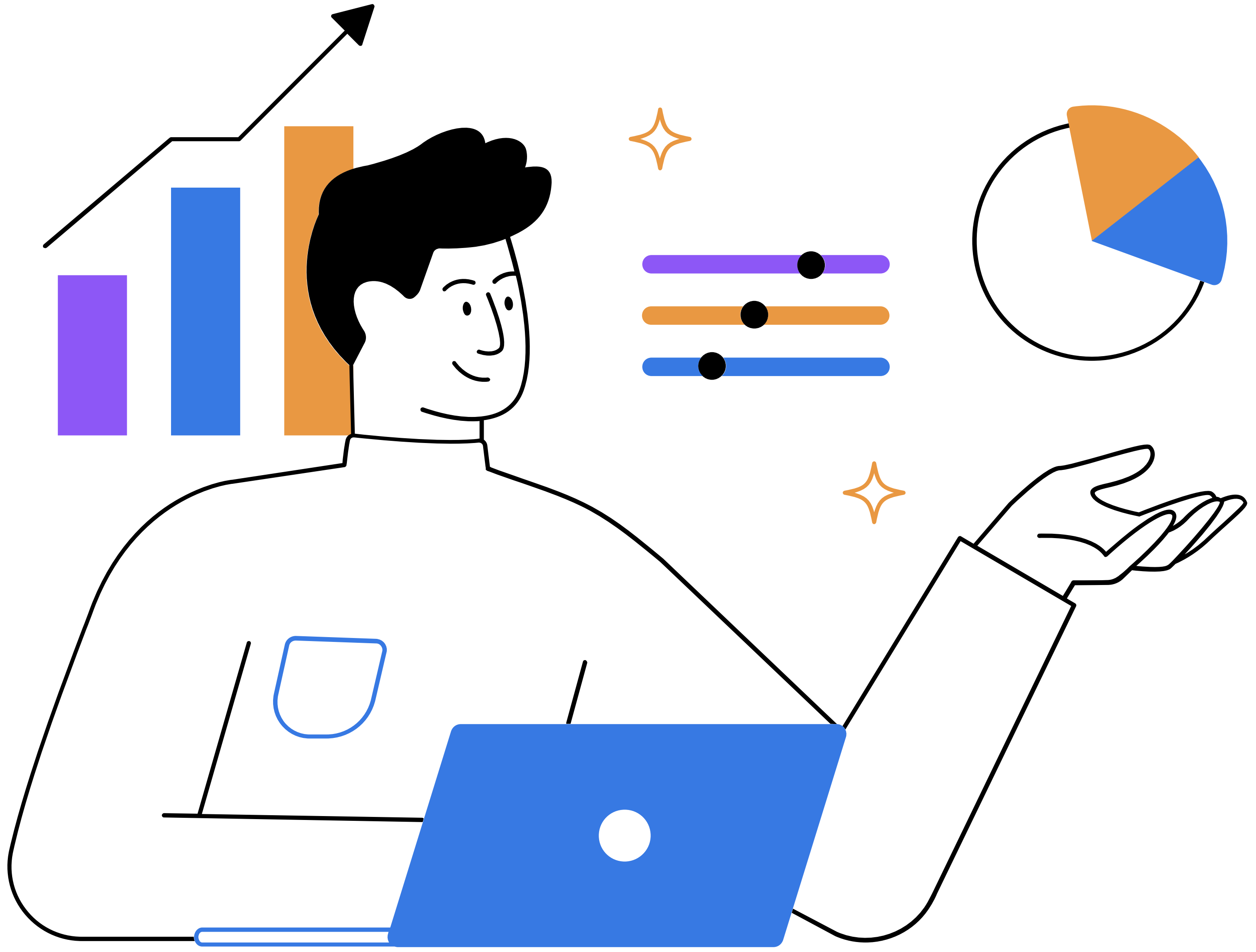


Discussion

Key challenges identified include data imbalance, short and ambiguous queries, and the absence of a direct link between search and click events. The system also remains sensitive to small variations in user behavior, such as typos or random clicks. Future work should focus on integrating deep learning models, improving feedback loops for retraining, and validating performance.

Conclusion

This study represents an initial step toward building a reliable, data-driven recommendation system for large-scale e-commerce environments. The proposed architecture integrates data ingestion, feature engineering, and predictive modeling within a flexible pipeline designed to handle complexity, scale, and dynamic user intent.



Although still conceptual, the project demonstrates a solid foundation for scalable and interpretable user behavior prediction. Once fully implemented, the system has the potential to enhance personalization on Best Buy's mobile site by delivering accurate, adaptive, and transparent product recommendations that align with evolving consumer interests.

Bibliography

[1] Kaggle, "ACM SF Chapter Hackathon – BIG Data," [Online]. Available:
https://www.kaggle.com/competitions/acm-sf-chapter-hackathon-big/data?select=test.csv

[2] Best Buy, "Official Website," [Online]. Available:
https://www.bestbuy.com/

[3] Repository for sources from workshops, "Workshop 1 – Data Mining Hackathon on BIG DATA (7GB) Best Buy Mobile Web Site," GitHub repository, [Online]. Available:
https://github.com/MelisaMelenge/Data-Mining-Hackathon-on-BIG-DATA-7GB-Best-Buy-mobile-web-site/blob/main/Workshop%201/Workshop1.pdf

[4] Repository for sources from workshops, "Workshop 2 – Design Phase," GitHub repository, [Online]. Available:
https://github.com/MelisaMelenge/Data-Mining-Hackathon-on-BIG-DATA-7GB-Best-Buy-mobile-web-site/blob/main/Workshop_2_Design/Workshop2.pdf

Thanks for your attention