E-commerce Data Analysis and Recommendation System

"Leveraging User Behavior for Enhanced Product Experiences"

S by Solomon Sannie

Azubi Africa

Project Overview: Unlocking User Insights

Our Core Goal

Analyze e-commerce data to deeply understand user behavior, predict critical item properties, identify abnormal user patterns, and develop sophisticated recommendation algorithms.

Driving Objectives

- Enhance the product page experience.
- Improve overall user engagement.
- Significantly increase conversion rates.

Our insights are powered by three key data sources: **events.csv** (user interactions), **item_properties.csv** (item attributes) and **category_tree.csv** (category hierarchy).

Unanswered Questions: Our Analytical Quest

Interaction Frequency

Which products or content are most frequently interacted with by users?

User Segmentation

Which user segments exhibit similar purchase or viewing behaviors?

Temporal Dynamics

How do seasonal trends affect product/content interactions, and when does purchase activity peak?

Conversion & Transactions

What is the visitor-to-purchase conversion rate, and what are the total transactions over time?

Methodology: Anomaly Detection

Purpose

To identify unusual user behaviors, which could indicate bots, fraudulent activity, or highly atypical user journeys.

Method

Utilizing the **Isolation Forest** model, a powerful unsupervised learning algorithm well-suited for outlier detection in high-dimensional datasets.

Key Features

Focused on aggregated user metrics: total number of events, average time spent per session, and unique item interactions per user.

Outcome

Successful identification of 7.63% abnormal users, enabling targeted interventions for security measures or customized user experiences.

Methodology: Item Property Prediction

Our objective was to infer specific item properties (e.g., categoryid, and available) within 'addtocart' events by leveraging known properties from preceding 'view' events for the same user or session. This enriches our understanding of the user's intent even when direct property data is absent.

Models Utilized:

- RandomForest Classifier
- LightGBM
- TNN (Tabular Neural Network)
- XGBoost

The models were trained on comprehensive visitor viewing behavior features, allowing us to predict the most likely item properties, thereby enhancing the dataset for more accurate recommendation systems.

Methodology: User Segmentation

> "Understanding our diverse user base is key to personalized engagement."

1

Purpose

To group users with similar interaction patterns and behavioral traits into distinct segments. This enables tailored marketing strategies and product adjustments.

2

Methodology

We employed **K-Means Clustering**, an unsupervised machine learning algorithm, which groups data points into a predefined number of clusters.

3

Features

Leveraged a rich set of user behavior features derived from event logs, including frequency of visits, types of interactions, and pages viewed.

4

Outcome

Identified clear user segments: Less Engaged, Moderately Engaged, and Highly Engaged. Each segment can now receive optimized content and offers.

Methodology: Recommendation System

Personalizing the E-commerce Journey

Our goal was to provide users with highly personalized and relevant item suggestions, driving both engagement and conversions.

Content-Based Filtering

Recommending items similar to those a user has liked in the past, based on item properties (e.g., category, brand, features).

Collaborative Filtering

Suggesting items based on user-item interactions, leveraging the preferences of similar users (user-user) or item similarities (item-item).

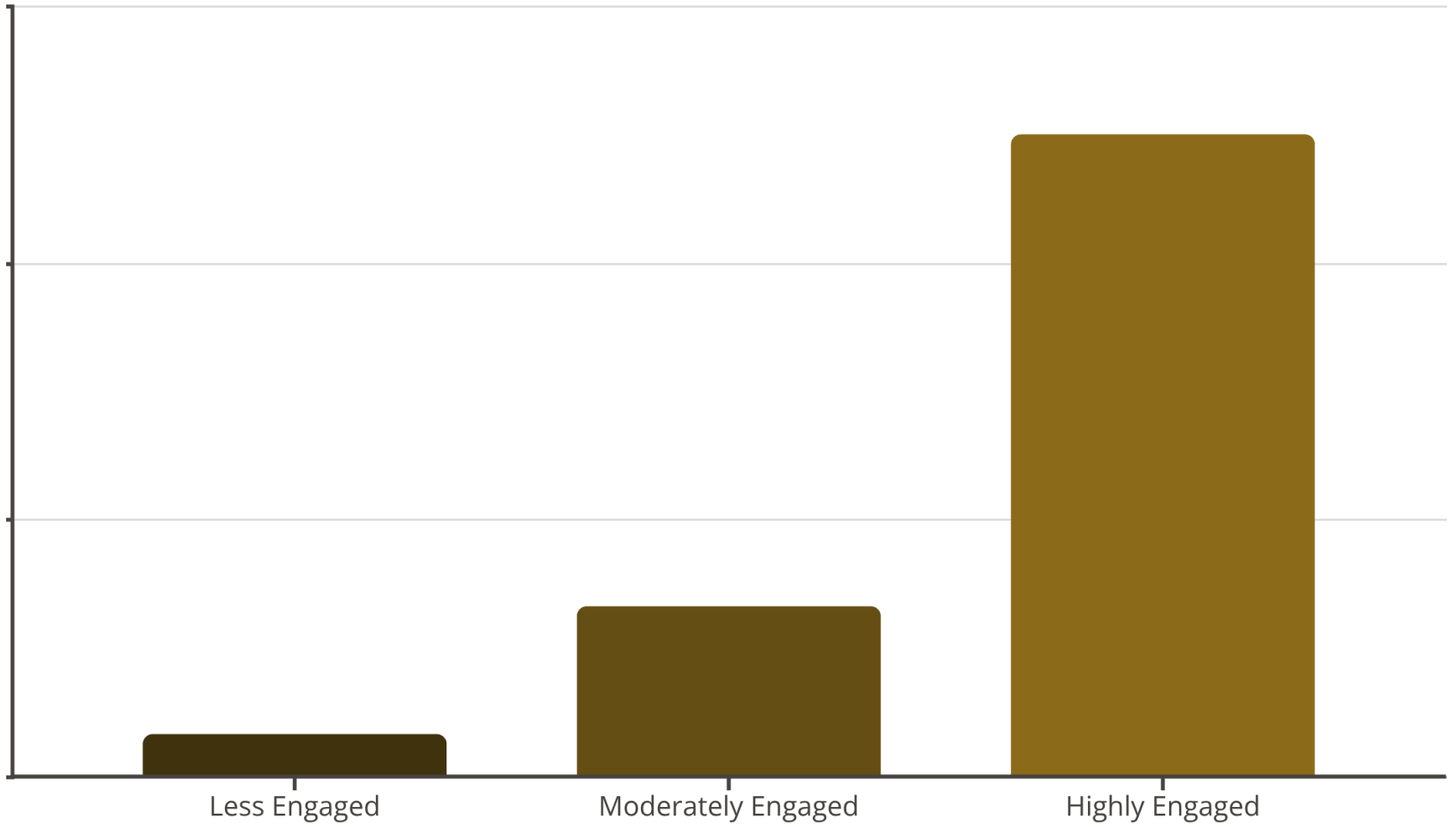
Hybrid Approach

Combining the strengths of both Content-Based and Collaborative Filtering to overcome limitations and enhance recommendation accuracy and diversity.

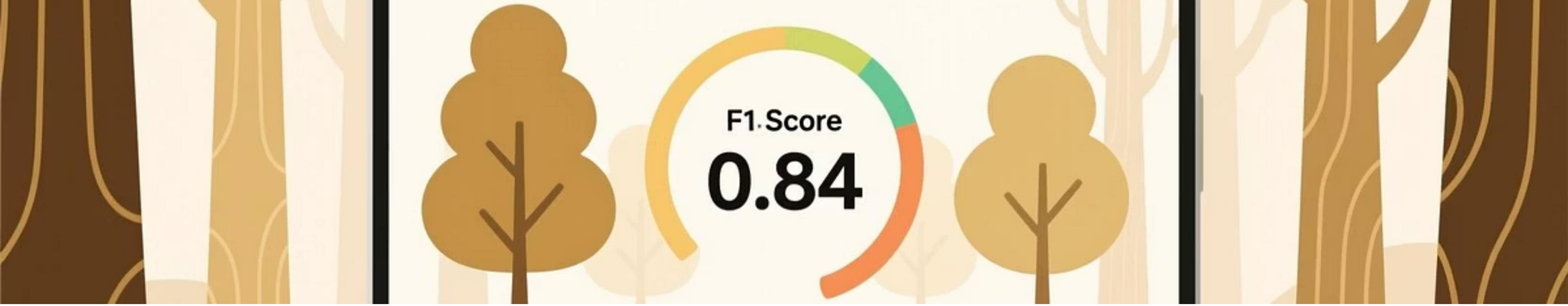
Category-Based Recommendations

Leveraging the category hierarchy to recommend broader or narrower selections of products, adaptable to user exploration patterns.

Key Results & Insights: User Behavior



- User Segments: Our clustering identified clear distinctions, from infrequent browsers to power users, enabling tailored interventions.
- Temporal Trends: Analysis revealed peak activity during evening hours (7 PM - 10 PM) and consistent high engagement on weekdays,
- suggesting prime times for promotions.
- Item Popularity: Key items and content frequently interacted with were identified, providing insights for inventory management and content strategy.



Key Results & Insights: Predictive Modeling

Anomaly Detection Success

Our Isolation Forest model effectively identified users exhibiting abnormal behavior patterns, ranging from bot-like activity to potential fraudulent interactions. This allows for proactive security measures and cleaner data for legitimate user analysis.

Item Property Prediction Robustness

The classification models, particularly RandomForest, demonstrated strong performance in predicting missing item properties from viewing data, achieving an impressive F1-score of around **0.84**. This validates our ability to enrich sparse datasets and improve recommendation accuracy.

EDA Insights: Purchase Behavior

7.63%

Abnormal Users

Only a small fraction of users in the sampled data was classified as abnormal in the system and was removed from the dataset for further analysis.

187946

Most Viewed Item ID

This product received the highest number of views across all sessions

0.83%

Low Conversion

The visitor-to-purchase conversion rate.

461686

Most Added to Cart

This item was most frequently added to shopping carts by users

EDA Insights: Predictive Analysis

320130

Most Predicted Item
ID

This item was most
frequently predicted.

Actionable Solutions: Enhancing Product Pages

Leveraging our analytical insights, we've identified key strategies to optimize product pages for improved user experience and conversion.

Personalized Recommendations

Implement our hybrid recommendation engine to dynamically suggest relevant products, increasing user engagement and potential purchases.

Segment-Specific Content

Customize product page layouts and content based on user segments, offering detailed information for highly engaged users and simplified views for less engaged ones.

Seasonal & Trend Highlights

Dynamically showcase items that are popular during the current season or trending periods, aligning with user interest and maximizing sales opportunities.

Anomaly-Aware Experience

Integrate insights from anomaly detection to tailor the experience for unusual user behavior, applying security checks or adjusted content to maintain platform integrity.

Actionable Solutions: Broader Impact

Beyond product page enhancements, our analytical framework offers widespread benefits across marketing, operations, and security.

Targeted Marketing Campaigns

Develop and deploy marketing strategies precisely crafted for identified user segments, enhancing conversion rates and customer loyalty through personalized content and offers.

Temporal Trend Optimization

Optimize website content delivery, promotional timings, and server load management by aligning with peak user activity periods, improving user experience and resource allocation.

Enhanced Security Protocols

Integrate anomaly detection insights into real-time security systems to proactively identify and mitigate fraudulent activities and bot attacks, ensuring platform integrity and user safety.

Next Steps

To fully realize the potential of our analysis and continue optimizing the e-commerce experience, we propose the following key actions:

Productionize Algorithms

Deploy the developed recommendation and tailoring logic into live systems, ensuring seamless integration with the e-commerce platform.

Conduct A/B Testing

Implement rigorous A/B tests to quantitatively measure the impact of new recommendations and tailored experiences on key business metrics like conversion rates and engagement.

Explore Advanced Modeling

Continuously investigate and integrate more sophisticated machine learning techniques to further enhance predictive accuracy and personalization capabilities.

Gather User Feedback

Establish continuous channels for collecting direct user feedback, which will be invaluable for iterative improvements and identifying new areas for enhancement.

Thank You

Questions?