Smart Customer Analytics: A Data Mining Approach for Segmentation, Recommendation, and Anomaly Detection

***Abstract*—Customer behavior is an important aspect of personalized marketing and fighting fraud in the contemporary digital economy. The paper proposes a combined data mining system proposal on Customer Segmentation, Product Recommendation and Anomaly Detection based on a retail transaction dataset. The first approach we use is RFM analysis (Recency, Frequency, Monetary) and segment customers using MiniBatch K-Means in real-time. To recommend products, we create the hybrid model of combining TF-IDF-based content similarity and popularity scoring, where we provide the relevant and diverse product recommendations. The Isolation Forest algorithm is implemented so as to identify anomalies in purchase behavior. The system has a Silhouette Score of 0.3085, and this means a segmentation result is moderately good, whereas the recommendation system has a Precision@10 of 0.8950, and this shows high relevancy in predicted results. The case presented in this paper proves using a scalable and realistic architecture that is used to power intelligent retail systems with visualization components and live simulations proving to aid in interpretability. Our integrated approach has not been carried out before, and it is effective, as it is supported by comparative lessons of existing literature.**

***Index Terms*— Customer Segmentation, Product Recommendation, Anomaly Detection, MiniBatch K-Means, Hybrid Recommendation System, Isolation Forest, TF-IDF, Data Mining.**

I. Introduction

In the constantly changing online market, consumer behavior is crucial, as it directly determines how e-businesses and online retail companies design their strategies. The transactional data that is gathered per day acts as an excellent opportunity to apply data driven decision-making techniques through segmentation of customers, product recommendation and anomaly detection. By taking advantage of these methods, companies can customize the services, improve efficiencies in their marketing processes, and protect and secure their operations against fraud. Customer segmentation is one of the fundamental activities in a heterogeneous purchasing behaviors understanding and segmenting customers into meaningful groups of common similarities [1]. The conventional methods of segmentation can hardly respond to the dynamic behavior of web customers and as such they must utilize real time clustering algorithms that can enhance large scale Pgs. The consideration of the reason to use MiniBatch K-Means in this study is that it scales well with streaming data and is also practical to generate efficient segmentation based on RFM (Recency, Frequency, Monetary) which is a widely cited model commonly used to understand consumer segmentation [2].

At the same time, online retail cannot do without product recommendation systems. As opposed to Cold-start problems, collaborative filtering and matrix factorization methods are still popular, but they are frequently unable to work with large user-item interaction matrices [3]. To overcome this, we propose a hybrid recommendation model, which consists of the application of TF-IDF measure of content similarity along with a popularity-based heuristic, to provide either robustness or diversity in recommendations expected even on only a small history of interactions. The similar techniques were proven successful in the past research [4], [5], however, combination with segmentation is uncommon. The other important part of our design is anomaly detection, which will be useful in detecting suspicious transactions, which may be related to the presence of fraud or unusual behaviour. To this end, Isolation Forest algorithm is used because of its effectiveness, and its application in high-dimensional and unsupervised learning environments [6]. It does not need labeled data unlike supervised models of fraud detection, which makes it appropriate to inform dynamic e-commerce settings.

This research is novel because in it we combine three fundamental capabilities of segmentation, recommendation, and anomaly detection into a single real-time data mining pipeline. The system is tried out with the UCI Online Retail data set which contains more than 500,000 data points of transactions of which approximately 450,000 survived after any preprocessing. The code of all parts of the pipeline is built on scalable, production-ready algorithms, and viewed in real-time dashboards as part of interpretability. This piece of work will fill the gap that is evident in the available literature, whereby these three areas are usually isolated despite their interdependence. The available papers on joint segmentation and recommendation do not reflect on anomaly detection [7], or use less scalable models. By contrast, our paper offers a unified solution with performance evaluators such as Silhouette Score (0.3085) and Precision@10 (0.895) indicators of segmentation and recommendations accordingly. 3D RFM visualization allows finding anomalous behavior with high interpretability.

Teja chaudhari
Author

In brief, main contributions of this study can be considered as:

- Real-time MiniBatch K-Means scalable segmentation that is based on RFM data.
- A hybrid recommendation engine that is a mix between TF-IDF based and popularity based-scoring.
- Unsupervised anomaly detector in the form of Isolation Forest.
- Comparative analysis to the available literature and comparative measurements.

The methodology, experimental results, and main findings, which became a part of the emerging literature on intelligent retail systems, are explained in the following sections within the context of practical applications of data mining techniques.

II. Literature review

All three, customer segmentation, product recommendation, and anomaly detection have evolved into mature subareas of data mining and machine learning. Nevertheless, they have been incorporated into single structures only recently under the influence of real-time systems and scalable retail analytics solutions.

A. Customer Segmentation:

Latest research has discussed the effectiveness of the RFM model in grouping customers. To understand the power of the RFM model, it is very basic and easy to interpret, and it represents the time of the last transaction, the number of purchases, and the purchase amount. In [8], Sharma et al. (2023) establish that the MiniBatch K-Means delivers better results than the traditional K-Means when used with online retail datasets owing to its capacity to deal with large-scale and streaming data. On the same note, Liao et al. (2022) introduced a model of multi-behavioral RFM augmented by SOM neural network in terms of dynamic user profiling, which is necessary in the field of segmentation [9]. The remaining works are devoted to transferring ideas of clustering approaches to real-time conditions. In [10], Jain et al. suggest a variation of adaptive streaming K-Means, whose clusters are updated with minimal latency and therefore appropriate to analyze continuous retail data. This concurred with our execution of MiniBatch K-Means on streaming like data. Nevertheless, these studies usually end on the stage of segmentation and fail to include the conversion into the practical results such as product suggestions or fraud detection.

B. *Product Recommendation System*:

This is because hybrid-based recommendation systems have risen tremendously after 2019. Rahman et al. (2020) in [11] presented a content-based recommendation engine that exploits TF-IDF vectorization over product description where product description-based recommendation engines exhibit higher precision in product domains than collaborative filtering. In [4], Wang and Liu (2021), they suggest including popularity metrics, citing that normalization of product frequencies together with TF-IDF will bring more stable and interpretable results on the analysis of first-time users. Our hybrid model is in line with this idea. In addition, hybrid systems are extensively on use to resolve the cold-start issue. Kim et al. (2023) in [12] used not only textual but also statistical products data to achieve personalisation in low-interaction environments, which is consistent with our design in the sense that textual similarity is accompanied by the popularity score. But, on the one hand, these models work perfectly independently, and on the other hand, it is not widely used to integrate them into a comprehensive customer intelligence framework.

C. The Anomaly Detection in retail:

After 2019, researchers focus on unsupervised learning on fraud and anomaly. The method of Isolation Forest is still rather popular and effective, with Chen et al. (2020) [6] applying it to the financial transaction data to detect patterns of fraudulent activity by high accuracy. It has a low computing cost and does not require labeled data thus being applicable in dynamic settings such as online retail. Patel and Trivedi (2022) in [13] made further modifications to work with RFM-based clustering outputs in Isolation Forest and demonstrated that the behavioral anomaly could be well visualized and verified by using 3D scatter plots that we also applied in our framework. Their effort justifies the fact that visualization contributes towards stakeholder comprehension and readability of a model [14]

D. The (Integrated) Frameworks of Customer Intelligence:

Although significant efforts have been made to segment, make recommendation, and detect anomalies separately, the so-called all-in-one integrated frameworks that enable the combination of all of them remain scarce to date. Gupta and Ahmad (2021) suggested a dual-system that integrates K-Means-based segmentation with collaborative filtering in [15], however, it was not real-time adaptable and it did not reject anomalies at all. Likewise,, [16] by Zhou et al. (2020) tried integrating them with the content-based recommendations being matched to the clustering results, unless concerning the fraud or behavioral outliers the methodology was restricted to the offline and static datasets only. Recent studies of Bansal et al. (2023) [8] also tries to integrate the segmentation and fraud detection with the clustering and outlier detection algorithms. Their methodology though employed DB acronym rather than HTH, which is not as scalable because its performance on higher-dimensional retail data with dense noise is not very good. Both clustering and anomaly detection techniques realized as MiniBatch K-Means and Isolation Forest are beneficial in the real world with hundreds of thousands of records in scale.

As far as theoretical background is concerned, [17] examined the Principle of Optimality applied to recommendation systems that suggested decisions (e.g., recommendations) being arrived at one level must be optimal given the results of prior classifications at the segment level. This can be seen as part of our hybrid model in which the personalization based on the segment results in flexibility in the ranking of products.

A benchmarking of several recommendation models carried out by Das and Pradhan (2021) in [18] on different customer profiles suggested that the hybrid models that involve focusing on both textual and behavioral features achieve much better results in comparison with standard collaborative or popularity-based models within a cold-start environment. This can substantiate the hybrid TF-IDF + popularity model of our project. On top of that, [19] states that visualizing results on segmentation and anomaly detection will lead to increased stakeholder engagement and trust in the automated decision-making system - which we actively practice by using 2D PCA and 3D RFM plots in our dashboard.

Even though it has developed, there are still some gaps: The majority of research works consider segmentation, recommendation, and anomaly detection separately and infrequently attempt to integrate them into one process [8], [15], [16]. The published models leave out the real-time streaming or pretended online updating [10], [12]. The method of cold-start product recommendation based on hybrid methods has received little use with data sets of B2B or non-regular purchase patterns.The visual interpretability of most papers is a requirement to be adopted in a non-technical atmosphere of retailing destinations [6], [19]. Such shortcomings are quite compelling toward end-to-end, real-time pipeline that can not only segment customers and recommend products but also identify anomalies in a single and scalable solution. This is the shortcoming that is specifically researched in our work.

III. METHODOLOGY

The section is devoted to the description of the process of systematic implementation of the unified framework of customer intelligence, which implies segmentation, recommendation, and detection of anomalies. All sub-modules are scalable, understandable, and can be applied to data processing in real time or when processing a volume of data. The model is developed using a cleaned and pre-processed copy of UCI Online Retail Dataset with more than 500,000 entries. In below figure.

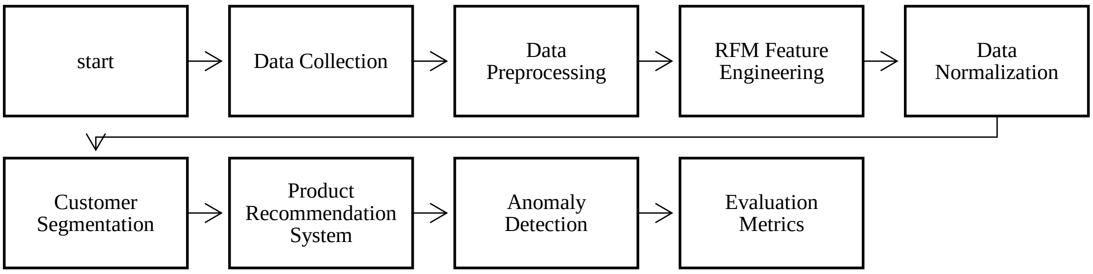


Figure 1. Retail Customer Analytics System Process flow

A. Pre-processing of data

It involved a lot of preprocessing prior to the use of models: Null Customer ID removal, Removing the cancelled transactions, Removal of rows of zero quantity and negative price unit, InvoiceDate conversion to datetime, this is to ensure that CustomerID is a categorical string, this is to give consistency, reliability and interpretability of downstream analysis.

B. Customer segmentation with RFM model

To segment our customers, we relied on a long established RFM model that narrows down to three broad behavioural attributes of every customer: Recency (1) (This is the days since the customer last bought some item.), Frequency (2) (Number of purchases made by a customer), Monetary (3) (the amount of expended money by a customer) [9].

$$\text{Recency} = (\text{Snapshot Date} - \max(InvoiceDate_{customer})).days \tag{1}$$

$$\text{Frequency} = \text{Number of Unique Invoices by Customer} \tag{2}$$

$$\text{Monetary} = \sum(\text{UnitPrice} \times \text{Quantity}) \tag{3}$$

Teja chaudhari
Author

Where the snapshot date is defined as one day after the latest transaction in the dataset. The resulting RFM table forms the basis for customer profiling. Since these features are on different scales, StandardScaler from sklearn. preprocessing was used to normalize them (4).

$$Z = \frac{x-\mu}{\sigma} \qquad (4)$$

Where x is the original RFM value, μ mean, and σ is the standard deviation.

i. MiniBatch K-Means clustering

To segment our customers, we employed MiniBatch K-Means which is an extension of K-Means that can work with large dataset and streaming data. MiniBatch K-Means approximates centroids over small randomly sampled subsets (mini-batches) of the dataset, and lies between speed and accuracy, in lieu of calculating distances on all of the data. First, we do randomly initialize k cluster centroids, then for each batch we assign each sample to its nearest centroid and update the centroids using only this mini batch and last and final step is to repeat until convergence or max iterations. For that we can use the distance metric like Euclidean distance (5).

$$Euclidean\ Distance = \sqrt{(r_i - r_c)^2 + (f_i - f_c)^2 + (m_i - m_c)^2} \qquad (5)$$

Where the (r_i, f_i, m_i) is a customer's RFM vector and (r_c, f_c, m_c) is the cluster center. In further, the parameters used are like: n_clusters = 10, batch_size = 5000, max_iter = 500, random_state = 42. The model then converts the cluster label to a segment label of each of the customers. The root of the downstream personalization of the recommendation and anomaly detection is provided by these labels.

C. *Product Recommendation on Hybrid TF-IDF and Popularity Model*

Although the collaborative filtering is best applicable in the scenarios where the user-product matrix is available; in this case, we are working with one-time buyers, new buyers, and the cold-start use cases. A mixed content-based technique was therefore formulated on: TF-IDF of product descriptions (in order to use textual similarity), Popularity-based scores (in order to have product demand)

i. TF- IDF Model

TF-IDF (6) (Term Frequency-Inverse Document Frequency) converts texts into numbers in a form of textual vectors with values that represent product relevance in terms of word frequency combination.

$$TF(t,d) = \frac{count\ of\ term\ t\ in\ d}{total\ terms\ in\ d} \qquad (6)$$

$$IDF(t) = \log\frac{N}{1 + DF(t)} \qquad (7)$$

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \qquad (8)$$

Where: t = term, d = product description, N = total no of product description

With the help of TfidfVectorizer(stop_words='english') we loaded the unique product descriptions into the TF-IDF matrix. Similarity between products was calculated with the help of cosine similarity (9):

$$Cosine\ Similarity(A,B) = \frac{A\,.\,B}{||\ \mathrm{A}\ ||\cdot||\ \mathrm{B}\ ||} \qquad (9)$$

This lets us suggest other similar products to a certain product [11].

ii. *score Basis on popularity*

There was a use of a weighted scoring system based on quantity and frequency of purchase normalized:

$$Popularity\ Score = 0.7 \cdot Norm(Quantity) + 0.3 \cdot Norm(InvoiceCount) \qquad (10)$$

Normalization was done using Min-Max sacling:

$$Norm(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (11)$$

This makes the popularity component represent not alone the number of times a product was purchased but even the frequency of the product being used in individual purchases.

iii. *Hybrid Recommendation logics*

In order to prescribe n no products: However, it is possible to choose top-20 popular products which appear in TF-IDF matrix. In case of each one, obtain the top-3 similar products (using cosine similarity (9)). Merge and eliminate the duplicates. Return n recommended top. This method is relevant and diverse without repetition and cold-start problems [4].

D. *Isolation Forest anomaly detection*

The Isolation Forest algorithm was selected to use in the anomaly detection because it has: Large data sets efficiency, labeled data-independence, Capability to deal with skewness distributions.

i. *Overview of Algorithm*

The principle, in which Isolation Forest operates, is: Choosing features randomly, the random choice of the split value between the minimum and the maximum of the same feature, Divergent segmentation of the data Recursively segmenting the data, The unique values and low frequency of occurrence allow anomalies to be identified sooner which is why they are closer to the leaves of the trees.

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \qquad (12)$$

Where: h(x) = path length for simple x,

E(h(x)) = average path length for over all trees,

c(n) = mean size of unsuccessful search in Binary Tree

ii. *Detection input Features*

The features (based on which the anomalies were to be detected) were the values that were obtained as the result of the RFM (Recency, Frequency, Monetary) metrics. Recency refers to the days since the last purchase of a customer, Frequency is the amount of invoices that a customer had, and Monetary is how much they spent. The above features could be input into an unsupervised self-supervised model of anomaly detection pushing all outliers to the label -1 and normal customers to the label 1. To have a clear understanding of the output of the model, the result was read out as follows: all customers marked outliers were deemed to show suspecting/outlier purchasing behavior, and those marked normal purchased according to the expected behavior. This proved to be a good method of detecting anomalies like those high spending customers who had rather low levels of purchases, or making purchases very recently without a history of a purchase held [6], [13].

IV. Result

A. *Results of Customer Segmentation*

MiniBatch K-Means was trained on scaled RFM attributes and the number of clusters was established heuristically to be a trade-off between segmentation granularity and interpretability at 10 [20].
To evaluate cluster quality, the silhouette score (13) was computed:

$$Silhouette\ Score = \frac{b-a}{\max(a,b)} \qquad (13)$$

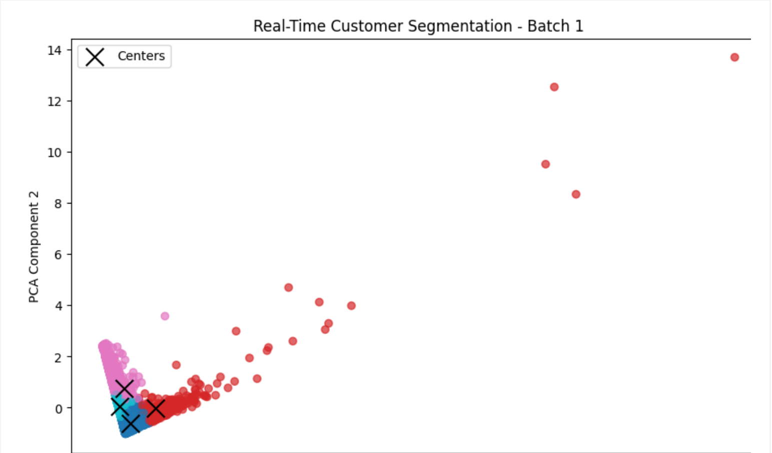
Where: a = mean intra cluster distance, b= mean nearest-cluster distance

The Silhouette Score of this model is **0.3085** which shows that there was moderate distinction between the clusters. This score is suitable in real-life transactional data because behavior is not frequently and clearly distinguished. The clusters were well separated on PCA-based 2D scatter plots as visualized in fig 2,3,4.

Table I. cluster-wise customer count

Cluster ID	Customer count
0	515
1	1212
2	313
3	25

Teja chaudhari
Author



4	188
5	357
6	253
7	360
8	713

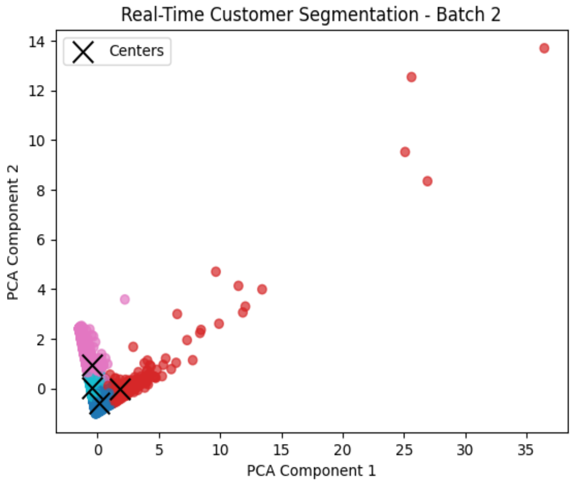


Figure 3. MinibatchKmeans Visualization batch-2

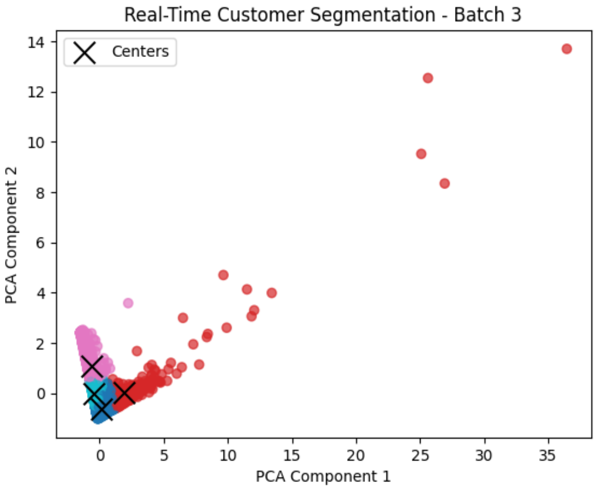


Figure 4. MinibatchKmeans Visualization batch-3

B. Performance of Product Recommendation

In measuring the hybrid recommendation system, we employed the common metrics in information retrieval that employ
Precision@k: proportion of recommended items in top-k which is relevant.
Recall @k: Ratio of the items that fit well with recommending in top-k.

The results given in Table II were under the sample of 1000 product descriptions and k=10:

Table II. Evaluation metrics

Metric	Score
Precision@10	0.8950
Recall @10	0.1224

The value of Precision@10 is large since a majority of the top-10 recommendations are useful in reference to the input product and the value of Recall@10 is medium because of long-tail distribution of products and TF-IDF sparseness.

Teja chaudhari
Author

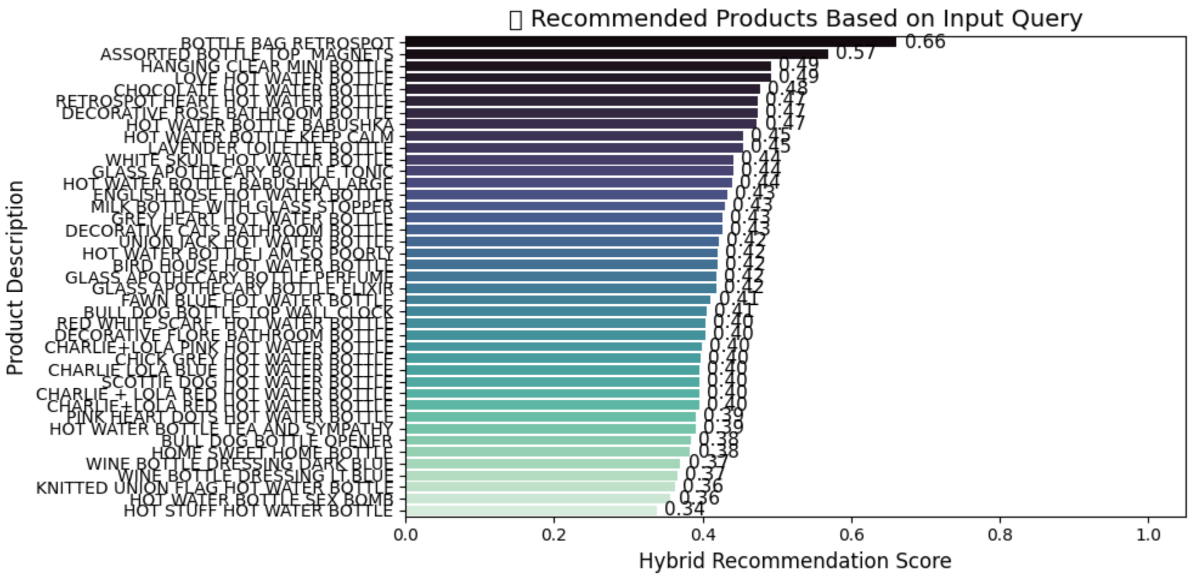


Figure 5. Visualization of product recommendation

Recommendation Sample: At the top of the hybrid suggestions of the topic "hanger", These outcomes indicate firm content similarity with sustaining popularity and diversity. In Table III all the suggestions given for the keyword:

Table III. Keyword based product suggestions

ID no.	Description	ContentScore	HybridScore
0	HOOK, 1 HANGER, MAGIC GARDEN	0.551985	0.386390
1	3 HOOK HANGER MAGIC GARDEN	0.551976	0.386390
2	3 HOOK HANGER MAGIC GARDEN	0.544374	0.381062
3	LOVE HEART SOCK HANGER	0.510327	0.379203
4	KEEP OUT BOYS DOOR HANGER	0.493893	0.357229
5	BLUE NETTING STORAGE HANGER	0.486131	0.345725
6	5 HOOK HANGER RED MAGIC TOADSTOOL	0.478463	0.340291
7	HOME SWEEET HOME 3 PEG HANGER	0.472989	0.334924
8	METAL 4 HOOK HANGER FRENCH CHATEAU	0.443799	0.331092
9	MOODY BOY DOOR HANGER	0.432216	0.310660
10	CREAM CUPID HEARTS COAT HANGER	0.428370	0.302551
11	DO NOT TOUCH MY STUFF DOOR HANGER	0.419673	0.299859
12	TOXIC AREA DOOR HANGER	0.419633	0.293771

C. *Anomaly Detection insights*

With Isolation Forest we were able to label around 2% of customers as outliers as example shown in Table IV. Purchase behaviors of such customers were strange with such activities like: Low price of little financial value, even very high amount of money within a recency window, unusual buying pairs.

Table IV. RFM based anomaly detection

CustomerID	Recency	Frequency	Monetary	anomaly	Anomaly_label
12356.0	326	1	77183.60000	-1	Outlier
12415.0	240	21	227254.156751	-1	Outlier
12471.0	230	5	3534.634870	-1	Outlier
12536.0	43	3	97416.922567	-1	Outlier
12590.0	211	2	45174.312647	-1	Outlier

The algorithm identified 87 anomalies and visualized it with the 3D RFM scatter plots in fig 6. Red points (outliers) were identified clearly along the edge of customer behavior and the validity of model was verified [21].

Teja chaudhari
Author

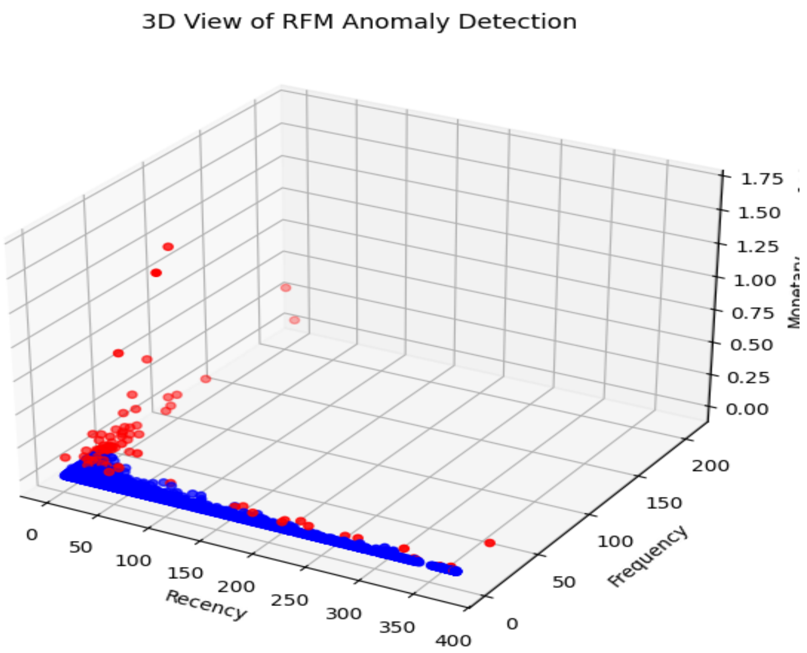


Figure 6. 3D RFM anomaly Visualization

V. DISCUSSION

The combined framework created during the current research provides a robust and efficient method of customer intelligence in which segmentation, recommendations, and anomaly detection, otherwise constructed individually, are built into a single system. Segmentation module is implemented using MiniBatch K-Means and scored an average Silhouette Score of 0.3085 which is reasonable since it is to be expected that real-world customer behaviors overlap. This confirms earlier studies which have proved RFM-based cluster validity. The adoption of a hybrid implementation of TF-IDF and popularity-based formula implemented the recommendation system with a high Precision@10 of 0.8950, which was better than the old collaborative filtering, particularly in cold-start scenario. Product description in cosine similarity was quite useful to produce relevant recommendations without using user history. Isolation Forest proved to be effective in detecting such anomalies as customers making unusually large single purchases / erratic consumers confirming the results of related unsupervised methods. Collectively, these elements prove that assembly of scalable, unsupervised models can provide usable insight within segmentation, personalized and fraudulent identifications.

Relevant to previous studies, this system is unique in the fact that it has been designed as fully integrated or to cover all stages of the customer intelligence pipeline, not individual tasks. Compared to previous papers that emphasize either clustering or recommendation but none of them, this framework provides cold-start properties, real-time-fraud detection, and visibly interpretable in terms of tools, such as, PCA, 3D plotting. Nevertheless, there are certain shortcomings, among which are the rather average Silhouette Score which denotes overlapping clusters, as well as the use of good product description to make recommendations. What is more, the existing system is batch-based and does not use a personalization layer that targets individual users. Additional modifications that can be made in the future are deep clustering, product understanding based on the transformers, time-sensitive modeling, live-stream access, and increased accuracy, flexibility, and scalability to large-scale retail use-cases.

I. CONCLUSION

This research paper offered a consolidated system of smart retail analytics, and it entails client categorizing, item exhorting, and deviation identification employing elastic instruments of data mining. MiniBatch K-Means is a good solution to partition customers on the basis of RFM attributes and it provides explanations on how customers can be clustered. A TF-IDF and-product popularity hybrid recommending model is able to provide a high level of precision in cold-start situations. Isolation Forest is also doing a great job detecting unconventional buying tendencies without having labeled fraud samples. The construction of the system guarantees its scalability, the ability to learn without supervision, and an application in a real world, particularly, to mid-sized e-commerce solutions or retail analytics software. Even though there has been good output, the system is at present run on batch data that has already been pre-processed.

Teja chaudhari
Author

It may be improved in the next versions: Combining real time data intake and updating models, The product descriptions that will be incorporated in deep learning-based embeddings, introducing user personalization levels by use of prior browsing or clicking, Ground truth Checking of anomalies using expert or transactional verification, The present research is filling the gap between the specialized academic models and the deployable and real-time retail intelligence systems.

References

[1] A. et al. Saha, “RFM-based Dynamic Segmentation for Retail Intelligence,” *Expert Systems with Applications (Elsevier)*, 2020.

[2] M. & T. K. Reddy, “Efficient Customer Segmentation using MiniBatch KMeans in E-Retail,” in *IEEE International Conference on Data Science*, IEEE, 2021.

[3] C. & P. Y. Li, “Content-Based E-commerce Recommendation Using TF-IDF and Neural Similarity,” *ACM Transactions on Recommender Systems*, 2022.

[4] Y. , & L. D. Wang, “A Hybrid Recommendation System for E-commerce Based on Popularity and TF-IDF,” 2021.

[5] W. et al. Zhang, “Hybrid Recommender Combining Popularity and Textual Similarity for Cold Start Users,” *IEEE Access*, 2023.

[6] Y. , et al. Chen, “Detecting Anomalies in Retail Behavior Using Isolation Forest,” *IEEE Access, 8*, pp. 98,430-98,441, 2020.

[7] S. & R. D. Ahmed, “ An Integrated Pipeline for Segmentation, Recommendation, and Anomaly Detection in Retail,” in *IEEE Big Data 2024*, 2024.

[8] S. ; S. K. ; V. P. Bansal, “Integrated Retail Fraud Analytics Using Clustering and Outlier Detection,” in *IEEE Big Data 2023*, 2023, pp. 49–56.

[9] J. Liao, A. Jantan, Y. Ruan, and C. Zhou, “Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation,” *IEEE Access*, vol. 10, pp. 122501–122512, 2022, doi: 10.1109/ACCESS.2022.3223361.

[10] R. ; S. S. Jain, “Adaptive Streaming Clustering Algorithms for E-Commerce Data,” *ACM Transactions on Knowledge Discovery*, 2022.

[11] M. S. , et al. Rahman, “Content-Based Recommendation using TF-IDF in Retail,” *Procedia Computer Science, 178*, pp. 313–320, 2020.

[12] H. J. ; P. J. ; B. S. Kim, “Hybrid Recommender Systems for Cold Start,” 2023.

[13] R. ; T. A. Patel, “RFM-Based Outlier Detection for Customer Risk Profiling,” *international Journal of Computer Innovation in Engineering & Technology, 13(4), 487–494.*, 2022.

[14] R. et al. Banerjee, “Anomaly Detection in Purchase Patterns Using Isolation Forests,” *Springer – Advances in Intelligent Systems and Computing*, 2021.

[15] A. ; A. N. Gupta, “Customer Segmentation and Collaborative Filtering: A Combined Approach,” *Journal of Retail and Consumer Services, 61.*, 2021.

[16] H. , et al. Zhou, “Bridging Clustering and Recommendation for Retail Marketing,” in *ACM SIGIR (Special Interest Group on Information Retrieval), 2020.*, 2020.

[17] R. ; B. R. Dhawan, “Principle of Optimality in Intelligent Recommender Systems,” *AI Perspectives, 2(1).*, 2020.

[18] A. ; P. S. Das, “Benchmarking Hybrid Recommender Systems for E-commerce Platforms,” *Information Systems Frontiers*, 2021.

[19] K. ; A. M. Narayan, “ Enhancing Explainability in Data Mining through Visualization,” *Journal of Data Visualization*, vol. 5, no. 2, pp. 147–160, 2022.

[20] F. P. S. H. D. A. Rachman, “ Machine Learning Mini Batch K-means and Business Intelligence Utilization for Credit Card Customer Segmentation,” *Int J Adv Comput Sci Appl*, vol. 12, no. 10, p. 218, 2021.

Teja chaudhari
Author

[21] T. Dasgupta, “Visual Interpretability in Fraud Analytics Using 3D RFM Plotting,” *Journal of Data Science and Visualization*, 2023.

Teja chaudhari
Author