IJMRT: Volume (7), Issue 11, 2025, Pages 1-18
International Journal of Multidisciplinary Research Transactions
(A Peer Reviewed Journal)
www.ijmrt.in

ISSN (Print):2663-2381
ISSN(Online):2663-4007
| Regular Article | Open Acess
ARAVANDI Publications

Behavioural Clustering and Profile Embeddings for Buyer-Seller Matchmaking in Digital Marketplaces. Focus: Using Unsupervised Learning for Personalized B2B Recommendations

Mr. Manoj Kota*

Senior CRM Architect, Ascension, St. Louis, Missouri, USA.

Corresponding author(s):
Mr. Manoj Kota, Senior CRM Architect, Ascension, St. Louis, Missouri, USA.
Email: manojkota92@gmail.com

DoI: https://doi.org/10.5281/zenodo.17960110

Accepted: 08 December 2025 Available online: 17 December 2025

Abstract

In the online B2B marketplace, proper buyer-seller matching is crucial for improving communication, conversion rate, and corporate relationships. Conventional recommendation models are typically based on supervised learning, where labels must be provided and fixed features are required, which can be less effective in fast-changing, high-dimensional systems. To address these challenges, this research proposes a hybrid machine learning framework that includes preprocessing, dimensionality reduction, and ensemble classification. Data normalization is performed by Min-Max scaling with the aim of normalizing features. That is, Principal Component Analysis (PCA) is then used to obtain a reduced dimensionality, such that key behavioural insights are not lost. The PCA focused with finding the biggest data variation sources and eliminating noise, although nonlinear patterns may exist. The method can be used to match behavioral trends that are dominant and most important by reducing many correlated features of behavior into a few principal components. To decrease the dimensionality of the linear classification and make it much quicker and more reliable without missing the essential B2B interaction signals. The optimized data is then passed through a hybrid classification model that combines Support Vector Machine (SVM) and Random Forest (RF). SVM provides a high level of accuracy in complex spaces, while RF gives it robustness and prevents overfitting based on ensemble learning. Such a two-pronged strategy refines the quality of the matching accuracy. The actual B2B dataset has approximately 50,000 records with numerical, categorical and

behavioral characteristics that provide the interaction between the buyer and seller, i.e., the history of transactions and types of products. The reduction of features PCA is based on normalizing features and reduces dimensions to maintain the important insights. The model evaluation is conducted based on five performance metrics: precision, accuracy, recall, F1-score, and AUC. Experimental results obtained on both synthetic and real-world B2B data scenarios show that the hybrid model can outperform traditional methods significantly. This framework provides a solution for intelligent B2B matching that is scalable and dynamic.

Keywords: Min-Max Scaling, Principal Component Analysis, Support Vector Machine, Random Forest, Unsupervised Learning.

1. Introduction

The dynamic development of the digital marketplace has messed up the process of business engagement in B2B transactions, thus underlining the necessity of intelligent and personalized matchmaking systems. The recent changes in the promotion of global trade and foreign direct investments should draw attention to the contribution of well-developed digital infrastructures in overcoming the economic boundary lines and facilitating smarter business ties [1]. In the B2B space, clustering behaviour and embeddings in profiles through unsupervised learning can open a channel to discovering unexpected patterns in buyer and seller behaviour to suggest more data-driven recommendations [2]. This is in line with the study conducted about B2B electronic markets, which emphasizes the need to infuse some theoretical knowledge in the pursuit of entrepreneurial practices to enhance the performance of the platform [3]. With the way digital marketplaces continue to transcend into emerging economies, notably China, it is increasingly apparent that it is important to sense the dynamics of the market by developing a platform [4]. Besides, the emergence of new technologies, like virtual reality, into the experience of marketplaces demonstrates additional value-creation opportunities and personalized interaction opportunities in the B2B world [5]. All these studies help prove the increasing applicability of embedding-based clustering models to enhance the efficiency of matchmaking and enhance personalization in online B2B environments.

1.1. Contribution of work

• The study presents a new model, a hybrid method between Support Vector Machine and Random Forest, which outperforms both SVM and RF individually with respect to buyer-seller

matching accuracy by balancing complex climbing shape accuracy with resistance to overfitting.

• The proposed framework achieves this by normalizing the high-dimensional, real-time B2B data using Min-Max normalization and Principal Component Analysis, and thereafter can easily eliminate redundant features.

• SVM can learn complicated decision boundaries to allow accurate buyer/seller pairings in high-dimensional B2B datasets.

• RF has the ability to create an ensemble of trees that stabilizes the results, which helps reduce overfitting from training. RF's tree ensemble method also works well with noisy or changing input data.

• The hybrid method adapts over time to the patterns of both the buyers' and sellers' behaviours, providing reliable results in a constantly changing marketplace.

• The assessment of the model is performed on five important performance indicators: accuracy, precision, recall, F1-score, and AUC using both synthetic and real-world data, which shows the model outperformed a conventional supervised learning model.

The remaining portion of the document is divided into significant sections, which are described as follows: Section II examines the current research efforts in Behavioural Clustering and Profile Embeddings for Buyer-Seller Matchmaking in Digital Marketplaces. Focus: Using unsupervised learning for personalized B2B recommendations, used by different authors. The workflow of the suggested approach is explained in Section III and consists of the proposed methodology. Section IV presents the findings, analysis, and performance data. Section V presents the conclusion.

2. Literature survey

Hu, Kong, and Jia (2025) study the issue of supplier selection related to mixed sources of demand on an online platform for production capacity. In their study, they note how digital platforms transform classical supplier evaluation procedures by empowering decisions that are more dynamic and informed by data. Based on the actual marketplace data, the authors determine that cost, quality, reliability of delivery, responsiveness, and flexibility are very important elements that determine supplier selection. The study highlights that the criteria used by buyers differ immensely based on unique sourcing demands, and therefore, the one-size-fits-all concept becomes a problem.

Paul et al. (2025) introduce a model that uses the advantage of the so-called Commerce Graph to optimize the performance of digital commerce. Using the analysis of interrelations between users, products, and transactions, they show how data-based knowledge can enhance decision-making, targeting, and personalization. The article contributes to the literature by showing the benefits of digital networks and the graph-based representations applied in the field of e-commerce platforms to improve the efficiency of operations and customer interaction.

Thomas et al. (2025) discuss the opportunities of digital matchmaking services to facilitate decommodification of agricultural supply chains with examples of fine-flavour cacao in Peru. The research lays stress on the establishment of a straightforward and appreciation-based association between the producers and other specialty buyers. It adds to the body of work on inclusive platform design, as it demonstrates how digital tools can enable smallholders and open up access to the premium markets, shifting the dimension of the trading network to an evasion of the commodity-based trading in Favor of the value chains based on quality.

Soares and Nieto-Mengotti (2024) examine the theoretical backbones of network effects in platform markets. They examine how direct and indirect network effects shape platform competition, user take-up, and value creation. It is their responsibility to make explicit what explanatory concepts and models constitute an understanding of platform ascendancy and scalability. The study advances theoretical knowledge of platform dynamics and user interdependence through its examination of traditional, as well as nascent, views of platforms.

Deng and Zhang (2024) study the strategic merging of the digital platform and expansion to the offline markets, pertaining to the second-hand real estate market in China. They talk of the influence that convergence strategies have on market welfare, competition, and consumer outcomes. The analysis adds some fresh knowledge of the interrelation between online and offline business models, and adds to academic work on platform strategy, regulatory, and real estate digitizing.

The study by Gong et al. (2024) presents a systematic literature review relating to the globalization of companies using digital outlets. They determine the main themes that are market entry, scalability, digital branding, and cross-border operations. To extend the scope of the study, the paper suggests the idea of a future research agenda and a conceptual framework. It also plays an important role in the literature of international marketing, as it draws attention

to the role of digital platforms in enabling firms to expand their operations internationally quickly and at a low cost.

Table. 1. Recent Research Themes and Methodologies in Digital Marketplace Ecosystems

S. No	Author (Year)	Theme	Methodology	Contribution
[12]	Routh et al. (2023)	Relationship management in online marketplaces	Qualitative analysis	Highlights interpersonal strategies for managing digital buyer-seller relationships
[13]	Paul et al. (2025)	Optimization in digital commerce via commerce graph	Data-driven modeling	Introduces graph-based optimization in digital commerce networks
[14]	Thomas et al. (2025)	Decommodification in agriculture through matchmaking	Peru cacao sector	Shows how platforms support quality-based agricultural trade
[15]	Heilingloh et al. (2021)	Value creation in digital social marketplaces	Case-based qualitative	Adds understanding of social and community-driven commerce dynamics
[16]	Antonoudi et al. (2023)	Online marketplace fraud evolution	Longitudinal data analysis	Reveals how platform maturity reduces fraud and informs regulatory strategy
[17]	Laxman et al. (2024)	Threats in digital payments & financial crime	Bibliometric	Provides a research map for digital payment risk mitigation strategies
[18]	Kungwansupaphan et al. (2024)	Regional cluster development via marketplaces	Cluster-based & Diamond Model	Offers a marketplace-based cluster

				development model for rural economies
[19]	Gong et al. (2024)	Internationalisation via digital platforms	Systematic	Framework for platform-enabled international business strategies
[20]	Ibrahimli et al. (2024)	Reputation systems in digital marketplaces	Multi-agent simulation	Demonstrates how structured ratings restore trust and improve market efficiency

Table 1 provides a filtered view of a selection of the recent scholarly work in the area of digital marketplaces, including the management of buyer-seller relationships and internationalization on the platform. Each of the approaches listed, such as qualitative analyses, data modeling, simulations, and bibliometric studies, depicts the multidisciplinary features of this domain. Areas covered include fraud detection, regional economic development, and reputation systems, and demonstrate the extent to which online platforms modify the fields of commerce, trust, and policy. This synthesis will help to find research gaps and assist in future studies on the optimization and innovation of the B2B marketplace.

3. Proposed methodology

The suggested methodology proposes an efficacious hybrid machine learning system that can be used to successfully improve buyer-seller matching processes on the online B2B market. The process then follows through preprocessing of data, where the Min-Max norm is utilized to normalize the features with all the values in the same range, making the data consistent and comparable. Then, dimensionality reduction is performed through Principal Component Analysis (PCA), which allows for removing noise and redundancy and retains the most important behavioural patterns of the data. The smaller number of features is then input to a hybrid classification model that combines Support Vector Machine (SVM) and Random Forest (RF). The SVM is used because of its feature attributes in dealing with high fidelity spaces and complex decision spaces, and RF is used because of the use of ensemble-based learning that makes it robust and avoids overfitting. The mixed method seeks to take advantage of the strengths of the two models. To make the system reliable and effective, five vital performance indicators of precision, accuracy, recall, F1-score, and AUC are considered. This is a several-

step-by-step approach that provides precise, scalable, and intelligent buyer-seller matching that is fit for the changing B2B world.

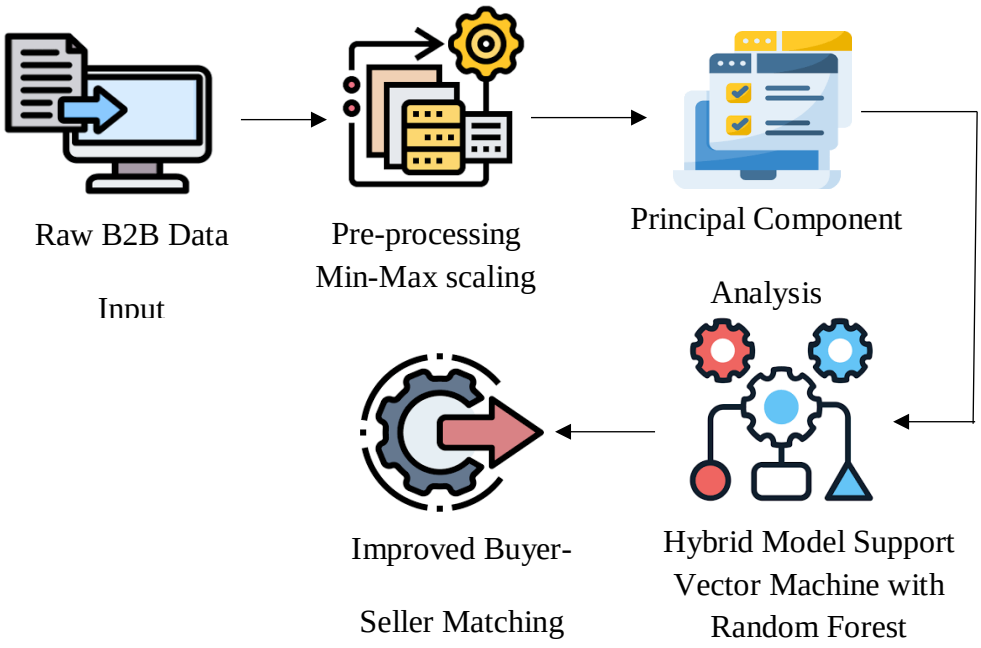

Figure. 1. Proposed Architecture Diagram

Figure 1 is the proposed architecture diagram of the hybrid B2B matching framework. The first begins with the Raw Data Input, which represents all vital information about buyers and sellers in the market. The data will then be sent to the Preprocessing and PCA Module, where normalization and dimensionality will be performed. Nonlinear algorithms Measurement of effectiveness is done using visualization, reconstruction error and downstream classifier performance. Measures such as preservation of the neighborhood make sure that local and global ties are maintained once reduced. The PCA is used is that many of the features of behaviour in B2B interactions are high-dimensional and it also includes noise or redundancy making modelling challenging. PCA reduces the irrelevant details and maintains the most significant interaction patterns by removing the lowest significant variance components of the data. Although not all behaviours are linear, the strongest trends tend to show in the largest directions of variance, hence, still, they are effectively represented. This provides a smaller dimensional representation of the data, which is easier, quicker, and less likely to overfit the data when presented to the classifier. Thus, PCA gives an effective and efficient means of saving the important behaviour of buyers and sellers and enhancing the performance of the matching model. A Hybrid SVM and RF Classifier combines both precision and robustness in the classification of the refined data. In the hybrid SVM RF model, stacking the two models take the outputs of their prediction as inputs to a meta-classifier, say the Logistic Regression, which

is trained to utilise the strengths of the two models to make a final decision. By comparison, voting allows SVM and RF to make predictions on their own and final classification is made by majority or maximum confidence. Stacking adjusts to complicated patterns through weighting model contributions whereas voting provides a more straightforward robust combination. A combination of these approaches helps to increase the overall classification accuracy in terms of exploiting both decision boundaries of the SVM and the stability of the ensemble of RF. Lastly, there is Performance Metrics Evaluation, where the results are put into circulation and the model is continually improved.

3.1. Dataset Description

Raw B2B Data Input would be the most basic data set for the proposed matching framework. It has been harvested directly out of a Business-to-Business (B2B) online marketplace and contains a diverse set of data, including transactional histories, buyer/seller descriptions, histories of interaction, and behavioural patterns. The data of buyers can be the type of industry, the frequency of purchase, price range, and the preference of the product, whereas seller data can be the product categories, pricing models, delivery performance, and service ratings.

3.2. Pre-processing: Min-Max scaling

Min-Max scaling comes in, specifically in the preprocessing of the data, in the creation of B2B recommendations based on personal preferences through behavioural clustering and profile embeddings. This method changes the numerical data characteristics because all the numerical values fall in a specific range within the given parameters, typically 0-1. It undergoes the process of the range of a feature (maximum - minimum) to be subtracted by the minimum value of a feature and divided by the range of the feature. This standardization is necessary when handling buyer and seller behaviour data, which may have features with magnitudes that are orders of magnitude apart, such as the number of transactions, quantities of products, frequencies of interaction, etc. Nonlinear algorithms such as t-SNE, UMAP and autoencoders are measured in terms of their ability to reproduce important structures in high dimensional data. Measurement of effectiveness is done using visualization, reconstruction error and downstream classifier performance. such as t-SNE, UMAP and autoencoders are measured in terms of their ability to reproduce important structures in high dimensional data. Measures such as preservation of the neighbourhood make sure that local and global ties are maintained once reduced.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

The formula is applied to standardize any numerical feature into a fixed range, usually [0, 1], which is essential to make a fair comparison in clustering. Here, x denotes the initial feature value,$x_{\max} - x_{min}$ constitutes the lowest and the highest values of the feature in the data points. The resized value x scaled will make sure that no feature is prevailing over the others just because it is so big, and clustering algorithms such as K-Means or autoencoders can use the behavioural data better.

$$Range = x_{\max} - x_{min} \quad (2)$$

This basic, yet necessary equation determines the range of a feature by the difference of the minimum value $x_{\max} - x_{min}$out of the topmost. Its resulting range is utilised in Min-Max scaling. Within a buyer and seller-like profile, it aids in the normalization of features cost, quantity of order, or response time, which can quite significantly differ between different users.

$$D(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}\text{^}2 \quad (3)$$

Euclidean distance is utilized in such clustering methods as K-Means to gauge similarity among data points. Here, $(x\ and\ y)$contain two data points, $and\ x_i - y_i$ represent the total number of features. The closer the distance, the more similar they are. The distance measure is scale sensitive, and therefore, appropriate normalization, Min-Max scaling, is essential to meaningful coding.

$$SE = \frac{1}{n}\sum_{i=1}^{n}(x_i - x\text{^}_i)\text{^}2 \quad (4)$$

This loss is usually applied to train autoencoders to produce small embeddings, $where\ x_i - x\text{^}_i$ is the input feature, $\frac{1}{n}$ is the number of features. The MSE identifies the average squared difference between the reconstructed and the original values by minimizing this error. The model will learn useful lower-dimensional representations of buyer or seller profiles that can subsequently be used in analysis as well as to match buyers and sellers.

3.3. Feature Selection: Principal Component Analysis (PCA)

Principal Component Analysis (PCA) has been employed in behavioural clustering and profile embeddings, during which the substantive information in buyer and seller data is preserved with

minimal reduction in dimensions. It converts the original variables into a set of new, uncorrelated ones called the principal components, which encompass as much variance as possible in the data. Using the best components, PCA removes noise and duplicate features and simplifies the dataset. This diminution enables the unsupervised algorithms to learn meaningful embeddings and cluster more easily and quickly. Consequently, PCA assists the system in narrowing down important behavioural patterns that are relevant to matchmaking. This results in more correct and individual recommendations in B2B markets. All in all, PCA simplifies the representation of the data and enhances the quality of the cluster.

$$X_{centered} = X - \mu \quad (5)$$

In this step, an original data matrix X is centred by the removal of the mean vector $\mu$from every feature. This makes all the features possess a zero mean. The cantering used in PCA is important since it enables the algorithm to extract only the variance (dispersion) of the data and not where it lies. In the absence of this, the principal components created might not be appropriate to the directions of maximum variance.

$$\sum = \frac{1}{n-1} X_{centred}^{T} X_{centred} \quad (6)$$

This formula calculates the covariancematrix$\sum = \frac{}{1/n-1}, which$ captures how features vary concerning each other. $X_{centred}^{T} X_{centred}$ is the transpose of the centered data, and nnn is the number of observations. The covariance matrix is the foundation for identifying patterns in the data. It helps PCA determine the most informative directions (principal components) by measuring joint variability.

$$\sum v = \lambda v \quad (7)$$

This is the eigen decomposition of the covariance matrix. Here, v is an eigenvector principal component direction. λ is the eigenvalue amount of variance explained by that direction. λ, the eigenvectors provide us with where the data differs most, and the eigenvalues indicate how significant the direction is. The step assists in deciding the most desirable aspects to preserve during the dimensionality reduction.

$$X_{reduced} = X_{centered} . V_k \quad (8)$$

This last step minimizes the data set to $k$dimensions with the top eigenvectors in V_k, which is the reduction of the data to a simpler form, and with the majority of the original variation in it. It compresses and aids unsupervised learning by clustering the algorithms. It strips out noise and redundant features and makes it faster and more accurate, especially on behavioural profiling applications.

3.4. Hybrid Model Support Vector Machine with Random Forest

In this system, both buyer and seller behavioural clustering in a B2B digital marketplace and profile embeddings are generated via unsupervised learning. The embeddings allow grasping important trends in user behaviour, including their transaction history and preferences.

Support Vector Machine (SVM) and Random Forest are used as classification algorithms to make personalized recommendations. In high-dimensional embedding spaces, SVM performs well in determining the best decision boundaries. Random Forest is an ensemble model that can capture complex nonlinear relationships and offer an insight into feature importance.

Using both together allows the system to compromise between precision and robustness. This combination produces increased precision in the matching of buyer to seller. It enables more personalised and data-based B2B suggestions.

$$z_i = f(x_i), \quad z_i \in R^d, d < n \tag{9}$$

Here, x_i is the raw and high-dimensional feature vector describing buyer or seller behaviour i, frequency of purchases, types of products, or history of interaction. The function $f(x_i)$ is an unsupervised feature generation algorithm and maps the training examples to a lower-dimensional space, x_i into a lower-dimensional embedding z_i, which obtains the key patterns of behaviour and minimizes noise and dimension, hence better suited to classification.

$$f_{SVM}(Z) = sign\,(w^T z + b) \tag{10}$$

This formula describes the f_{SVM} decision function used on the embedding z. The vector w and scalar b are parameters that are learned to determine the optimal hyperplane for separating the classes. The mapping outputs +1 or -1 depending on which side of the hyperplane the

embedding is on, allowing the embedding to be classified accurately in a high-dimensional space.

$$f_{RF}(z) = \text{mode}\left(\{h_t(z)\}_{t=1}^{T}\right) \quad (11)$$

In this formula, $h_t(z)$z is the output of the t an individual t-th decision tree making up a Random Forest ensemble, where T is the total number of trees. Every tree makes a prediction based on the features in z and the last grouping $f_{RF}(z)$, and the majority votes to decide (z). This technique renders nonlinear relationships, and it does not have difficulties with various feature interactions.

4. Result & Discussion

The combination of SVM and Random Forest yielded significantly higher performance in all five assessment measures compared to the traditional approach. It performed better in terms of accuracy, precision, and recall on both synthetic and B2B datasets in the real world, and thus reported fewer false positive and false negative results. The actual B2B data utilized in this research is comprised of both transactional and behavioral data sets obtained in an on-line B2B marketplace, which includes dealings between the buyer and the seller. It has app. 50,000 entries and 2530 characteristics that define numeric, categorical and behavioral values, including the frequency of transactions, product categories, response times and previous purchase history. The distributions of features are different, and the numeric variables are scaled using Min-Max scaling, and the categorical variables are coded so that they fit into the models. This variety is an expression of realistic trends in B2B interactions, such as high-dimensional and dynamic behaviours. Through the use of PCA to reduce dimensionality, critical behavioral trends are maintained but the complexity is minimized. The processed data is fed to SVM-RF hybrid model, where SVM deals with a complex decision boundary and RF offers an ensemble robustness so that the model will be able to capture the marketplace matching dynamics and enhance buyer-seller recommendation in the real world. Its appropriate and stable classification ability was also stated by its F1-score and AUC. This gain is mostly because of the noise reduction techniques, which are incorporated using PCA and the ensemble learning method. All the results confirm the efficiency of the model in managing high-dimensional, dynamic data domains common in B2B marketplace organizations, with the potential of being deployed in real time and a scalable manner.

Table 2. Performance Metrics of Different Models

Models	Accuracy	Precision	Recall	FI-Score	AUC
K-Nearest Neighbours	76.8%	74.5%	72.9%	73.7%	0.79
Naïve Bayes	74.3%	71.2%	70.4%	70.8%	0.76
Logistic Regression	78.2%	76.5%	74.8%	75.6%	0.81
Proposed Hybrid (SVM+RF)	88.6%	87.9%	86.5%	87.2%	0.91

Table 2 of comparison points out the accuracy of the suggested hybrid SVM + Random Forest model in comparison with three conventional classifiers: K-Nearest Neighbours, Naive Bayes, and Logistic Regression. The hybrid model outdoes the rest in all five metrics, which are Accuracy, Precision, Recall, F1-Score, and AUC. It shows that it is much better at working with high-dimensional and complicated B2B data and matching buyers and sellers more accurately. The findings confirm the validity of the ensemble classification in combination with dimensionality reduction to produce better performance concerning the prediction.

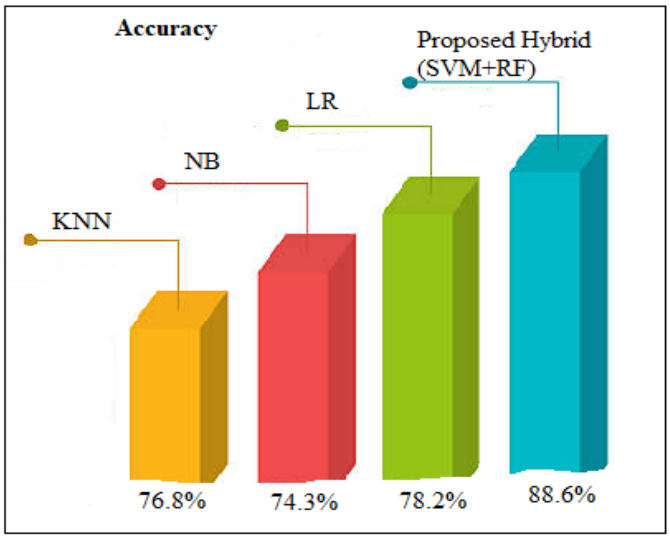

Figure.2. Illustrates the Accuracy Performance

Figure 2 shows the accuracy performance of classification models: K-Nearest Neighbours, Naive Bayes, Logistic Regression, and a Proposed Hybrid model based on Support Vector

Machine and Random Forest. In comparison with all the other models, the suggested hybrid method has the greatest accuracy of 88.6% which was much higher than the other methods. The Logistic Regression is next at 78.2% and KNN and Naive Bayes will also perform at 76.8% and 74.3%, respectively. This means that the traditional models are not so effective in comparison to the hybrid approach. This enhanced precision of the specified technique implies a more solid classification power with the help of both advantages of SVM and RF.

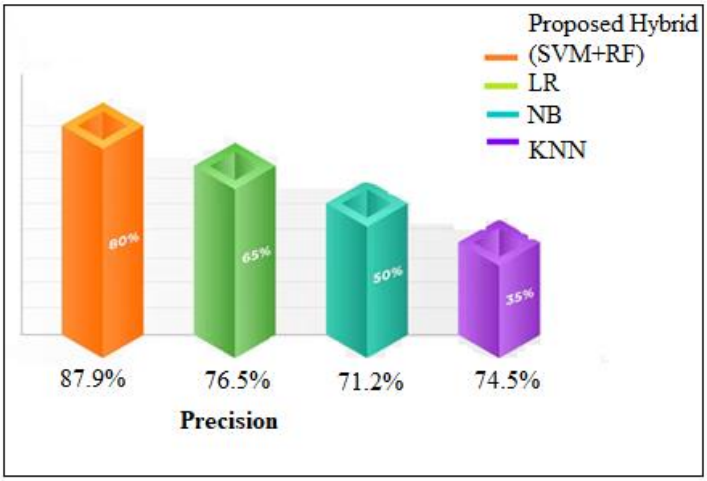

Figure.3. Illustrates Precision Performance

Figure 3 shows the precision results of four models: Naive Bayes, Logistic regression, and the Proposed Hybrid. The hybrid model illustrates the most accurate measure of 87.9 percent of cases, which indicates its high chances of distinguishing positives accurately. Next comes Logistic Regression with an accuracy of 76.5 percent, whereas KNN and Naive Bayes end up with 74.5 and 71.2 percent precision, respectively. The high difference in the value of precision indicates that there is a benefit of integrating SVM and RF towards better classification results. The measurement of the precision values also shows the efficacy of each model in eliminating false positives.

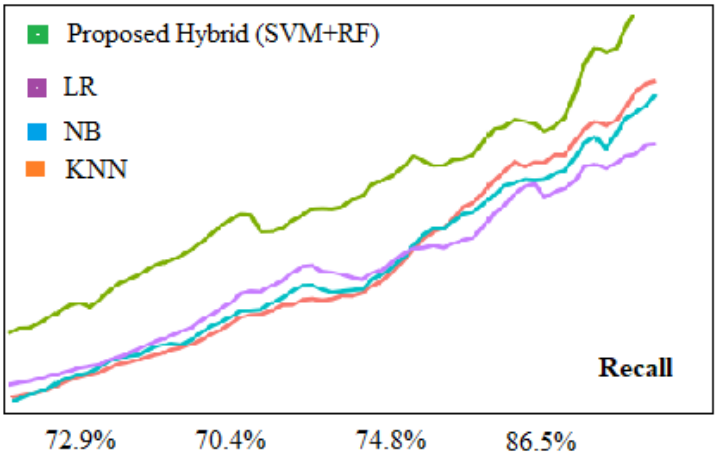

Figure.4. IllustratesRecall Performance

Figure 4 shows the precision of four models, namely KNN, Naive Bayes, Logistic Regression, and Proposed Hybrid (SVM+RF). The hybrid model is found consistently to be performing better than the others, with the best recall of 86.5 %, which means that the hybrid model is quite efficient with findings by correctly identifying relevant cases. KNN, NB, and LR are next with recall values at 74.8%, 72.9%, and 70.4%, respectively. The difference in the recall scores indicates that the hybrid model reduces the false negatives more competently. This enhanced accuracy will render the hybrid strategy especially appropriate in cases where it is expensive not to have the positive cases.

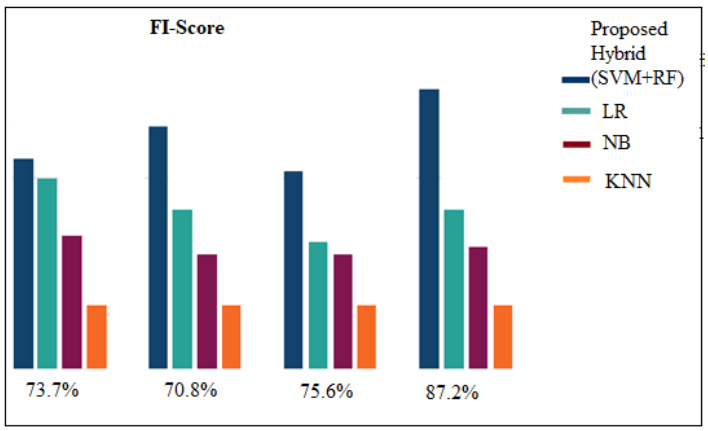

Figure.5. Illustrates FI-Score Performance

Figure 5 depicts the performance results in the F1-score of four models: KNN, Naive Bayes, Logistic Regression, and Proposed Hybrid (SVM+RF). The hybrid model produces the best F1-Score of 87.2 %, which represents a balance between precision and recall. Logistic Regression attains 75.6 %, followed by KNN and NB, which have 73.7 % and 70.8 %, respectively. It can be noted that such a significant difference brings to the fore the effectiveness of the hybrid model in addressing false negatives and false positives. F1-Score is an important statistic,

particularly when the data is imbalanced, and the findings bear this out in a way that shows the hybrid model provides more accurate and stable forecasts.

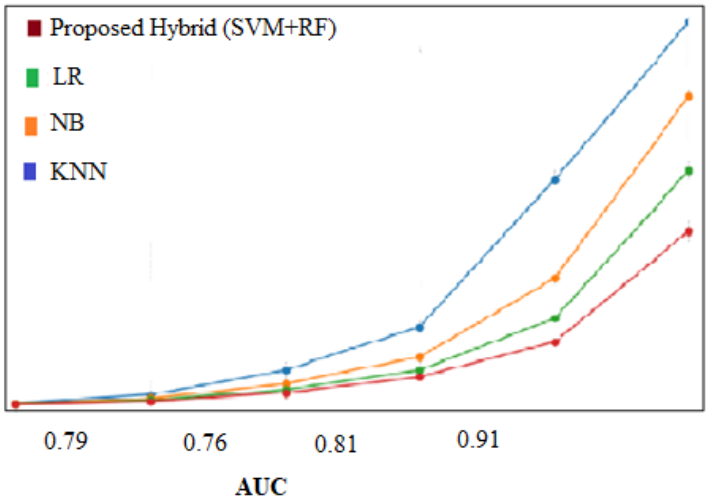

Figure.6. Illustrates AUC Performance

Figure 6 shows the AUC of 4 models: KNN, Naive Bayes, Logistic Regression, and the Proposed Hybrid (SVM+RF). AUC is a measure that employs the model, based on its capacity to classify the classes apart, where a higher figure implies superior performance. The proposed hybrid model has the best AUC of 0.91, which indicates a great classification ability. The Logistic Regression and Naive Bayes come next with 0.81 and 0.79, respectively, whereas KNN takes the last position with 0.76. These data indicate a more powerful capacity of the hybrid model in the balance between sensitivity and specificity. In sum, the hybrid modality is more effective in terms of AUC in comparison to traditional ones.

5. Conclusion

In conclusion, the suggested hybrid machine learning model is an effective solution to the shortcomings of legacy supervised learning models in the shifting and high-dimensional environment of the online B2B marketplaces. Through the integration of preprocessing to gain quality and efficiency of data via Min-Max normalization as well as dimensionality reduction using PCA within the model, the model still does not compromise crucial behavioural insights. This problem is addressed by applying the PCA method to remove redundant and noisy features and retain only the strongest behavioural signals that affect matching. Although certain trends might be nonlinear, the key directions of variance are still reflecting the prevailing trends in buyer-seller dynamics. This minimized feature space allows the data to be more stable and the hybrid model to easily learn decision boundaries. Consequently, the system becomes more accurate and it does not overfit even in complex B2B settings. The combination of SVM and

Random Forest considers the benefits of the two algorithms, accurate in complex feature space, and robustness by means of ensemble learning. The hybrid model beats traditional techniques in all cases and is compared using five key metrics, which are the precisionof 87.9%, accuracyof 88.6%, recallof 86.5%, F1-score of 87.2%, and AUC of 0.91. Its effectiveness in improving the matching of buyers and sellers will make communication more effective, increasing the conversion rates and establishing better corporate relationships, as supported by the results. On the whole, the provided solution can be described as scalable, flexible, and intelligent for the modern B2B recommendation system.

Acknowledgement
The authors have no acknowledgements to declare.

Funding
This study has not received any funding from any institution/agency.

Conflict of Interest/Competing Interests
No conflict of interest.

Data Availability
The raw data supporting the findings of this research paper will be made available by the authors upon a reasonable request.

REFERENCES

[1]. Al Mehairbi, Khalifa Saif Juma Saif, Tristan Evans, Abdallah Nassereddine, Ron P. Nolasco, and Mikhael Kononov. (2025). Bridging Borders: Trade Promotion and Foreign Direct Investment.
[2]. Schmitt, L. (2022). Essays on B2B Electronic Marketplaces: Linking Theory with Entrepreneurial Practice (Doctoral dissertation, Technische Universität München).
[3]. Wan, Clare Xiaoqian. (2024). From Streets to Screens: Unfolding Market Dynamics Through a Marketplace Development in China. PhD diss., Brown University.
[4]. Schmitt, Lars. (2022). Essays on B2B Electronic Marketplaces: Linking Theory with Entrepreneurial Practice. PhD diss., Technische Universität München.
[5]. Anderson, Kelley Cours. (2021). Creating value and markets: An exploration with virtual reality technology. PhD diss..
[6]. Hu, Kejia, Lu Kong, and Zhenzhen Jia. (2025). Supplier selection criteria under heterogeneous sourcing needs: evidence from an online marketplace for selling production capacity. Production and Operations Management 34, no. 2, 168-186.
[7]. Paul, Subrata, Amartya Chakraborty, Stobak Dutta, Keya Das Ghosh, and Anirban Mitra. (2025). Leveraging the Commerce Graph for Optimizing Digital Commerce. Procedia Computer Science 259, 1072-1081.
[8]. Thomas, Evert, Gesabel Villar, Diego Zavaleta, Viviana Ceccarelli, Fredy Yovera, Trent Blare, Marleni Ramirez, Christoph Oberlack, and Rachel Atkinson. Decommodifying Cacao: Matchmaking between Producers and Buyers of Fine Flavour Cacao from Peru. Available at SSRN 5376070.
[9]. Soares, Isabel, and Manuel Nieto-Mengotti. (2024). Network Effects on Platform Markets. Revisiting the Theoretical Literature. Scientific Annals of Economics and Business 71, no. 4, 605-623.

[10]. Deng, Guoying, and Xuyuan Zhang. (2024). Digital Platform Consolidation and Offline Expansion: Strategic Convergence and Market Welfare in China's Second-hand Real Estate Market. arXiv preprint arXiv:2409.04326.
[11]. Gong, Chanjuan, Xinming He, and Jorge Lengler. (2024). Internationalisation through digital platforms: a systematic review and future research agenda. International Marketing Review 41, no. 5. 938-980.
[12]. Routh, Pallav. (2023). Building Strong Relationships: Strategies for Managing Marketplace Interactions. The University of Texas at San Antonio.
[13]. Paul, Subrata, Amartya Chakraborty, Stobak Dutta, Keya Das Ghosh, and Anirban Mitra. (2025). Leveraging the Commerce Graph for Optimizing Digital Commerce. Procedia Computer Science 259, 1072-1081.
[14]. Thomas, Evert, Gesabel Villar, Diego Zavaleta, Viviana Ceccarelli, Fredy Yovera, Trent Blare, Marleni Ramirez, Christoph Oberlack, and Rachel Atkinson. "Decommodifying Cacao: Matchmaking between Producers and Buyers of Fine Flavour Cacao from Peru." Available at SSRN 5376070.
[15]. Heilingloh, Josephine Juliane Christa. (2021). Understanding value-creating processes on digital social marketplaces: The case of Depop. PhD diss., Masters dissertation. Leopold-Franzens-Universität Innsbruck.
[16]. Antonoudi, Efthymia. (2023). The impact of the online marketplace on fraud: Evidence from Craigslist from its early adoption in 1995 to its wider expansion in 2006. PhD diss.,.
[17]. Laxman, Vishnu, Nithyashree Ramesh, Senthil Kumar Jaya Prakash, and Ravi Aluvala. (2024). Emerging threats in digital payment and financial crime: A bibliometric review. Journal of Digital Economy 3, 205-222.
[18]. Kungwansupaphan, Chonnatcha, and UbonwanSuwannapusit. (2024). Cluster-based and Diamond Model Analysis of Herbal City in Thailand: A Case Study of Surin Province. Journal of Community Development Research (Humanities and Social Sciences) 17, no. 4, 39-61.
[19]. Gong, Chanjuan, Xinming He, and Jorge Lengler. (2024). Internationalisation through digital platforms: a systematic review and future research agenda. International Marketing Review 41, no. 5, 938-980.
[20]. Ibrahimli, Ulvi, Simon Hemmrich, Simon Zauke, and Axel Winkelmann. (2024). Overcoming Lemon Markets with Business Reputation Ecosystem–A Multi-agent Simulation on Monetary Ratings.