Anticipating User Desires: Predicting Software Product Feature Demand through Consumer Behavioral Analytics

Abdullah A. Aldaeej1

1Department of Management Information Systems, Imam Abdulrahman Bin Faisal University, Dammam 34221, Saudi Arabia.

Correspondence should be addressed to Abdullah A. Aldaeej; aaaldaeej@iau.edu.sa

## Abstract:

In this study, we introduce the Dynamic Behavioral Feature Predictor (DBFP), a machine learning model designed to predict feature demands for software products based on consumer behavioral analytics. The model leverages advanced techniques such as Deep Convolutional Neural Networks (1D CNN) and Long Short-Term Memory (LSTM) networks to identify complex user behavior patterns. We demonstrate that DBFP achieves a high accuracy of 99%, outperforming established models such as LSTM and 1D CNN in extensive comparative experiments. Although the model shows promising performance, we note that its accuracy may be sensitive to data quality, and further research is needed to evaluate its robustness in real-world applications. DBFP excels at identifying diverse user behavior types, offering tailored recommendations for software features that align with individual user preferences. This work highlights the potential of behavioral analytics in personalizing software development, enhancing the user experience, and improving the efficiency of feature prioritization. By bridging the fields of machine learning, big data (BD), and software personalization, DBFP lays the foundation for future advancements in user-centric software engineering. This study not only contributes to the field of predictive analytics but also opens new avenues for applying behavioral insights to software design and development.

**Keywords:** Consumer Behavioral Analytics, Software Product Development, Predictive Analytics, Sequential Pattern Mining

# Introduction

Consumer behavioral analytics have significantly transformed the landscape of software product development by offering previously unseen insights into user preferences and behaviors. This research seeks to explore the intricate relationship between user actions and the features demanded by software products, with the aim of better anticipating user needs through predictive analytics [1]. By leveraging comprehensive datasets derived from user interactions and engagement patterns, this study aims to equip software development teams with the tools to uncover latent user desires and preferences. Central to this research is the application of advanced analytical techniques and machine learning algorithms, which have the potential to identify hidden patterns in complex interaction data. These tools empower organizations to make more informed decisions, enhancing their ability to prioritize feature development efforts effectively. By bridging theoretical predictive analytics and practical application, this study intends to support software development teams in understanding and addressing evolving user demands [2]. Motivating this research is the recognition that traditional methods of feature prioritization often fail to capture the nuanced and dynamic preferences of modern consumers [6]. Conventional approaches typically rely on subjective evaluations or historical data, which may no longer be sufficient in predicting user behavior. In contrast, our approach leverages consumer behavioral data to reveal hidden trends and offer a more precise understanding of user preferences. This shift towards data-driven strategies is intended to help organizations develop more responsive, user-centric products, resulting in improved user engagement and retention. Furthermore, this study is driven by the desire to surpass outdated product design paradigms and provide a user experience that truly meets contemporary expectations. By anticipating user needs, software development teams can create products that align more closely with user preferences, ultimately enhancing product success. The research also addresses a crucial gap in connecting predictive analytics theory with its real-world applications. Many organizations struggle to implement these theoretical concepts into their development processes [11]. Through empirical validation, this research aims to provide practical insights that software teams can integrate into their workflows to effectively incorporate predictive analytics. Despite the promising potential of predictive analytics, there are significant challenges to overcome, such as concerns about data privacy, the complexity of algorithms, and a lack of organizational readiness [23]. For predictive analytics to achieve its full potential in driving innovation and fostering user-centric product design, these obstacles must be addressed. This research ultimately seeks to promote a shift towards a more agile, user-centered approach to software product development, laying the groundwork for a future where software design is driven by data insights that accurately reflect user demands and preferences.

Effective demand forecasting is crucial for optimizing inventory management, production planning, and resource allocation, leading to improved operational efficiency and profitability for organizations. By accurately predicting demand, businesses can identify future opportunities, mitigate potential risks, and adapt to shifts in customer preferences.

In light of this, we have defined the following research objectives to address the challenges and deepen our understanding of the subject:

1. Develop a robust model that predicts user preferences for new software features by utilizing consumer behavioral analytics.
2. Examine complex interaction data to evaluate how well various machine learning algorithms can uncover hidden user demands and patterns.
3. Explore innovative techniques, such as sequential pattern mining and sentiment analysis, to extract valuable insights from diverse customer behaviour datasets.
4. Assess the impact of predictive analytics on software product development, including feature prioritization and budget allocation, with a focus on user satisfaction.

Through these objectives, this study aims to enhance the prediction capabilities of software development teams, ensuring that new features align closely with evolving user expectations. This will foster a more flexible and user-centric approach to software design.

By addressing key issues and expanding our understanding of software development and predictive analytics, this research makes a significant contribution to the field. Specifically, it seeks to:

1. Create a transformative predictive model that integrates consumer behavioral analytics to accurately forecast user preferences for upcoming software features.
2. Advance knowledge on the effectiveness of different machine learning algorithms in identifying latent user demands and patterns from complex interaction data.
3. Introduce innovative methodologies for extracting meaningful insights from various consumer behavioral data sources, including sentiment analysis and sequential pattern mining.
4. Evaluate the influence of predictive analytics on user satisfaction and software development processes, helping guide feature prioritization and resource allocation.

By accomplishing these goals, the study will enhance software development teams' ability to make accurate predictions, enabling them to create more user-centric products that meet changing demands.

The paper is structured into five key sections. The Introduction outlines the research problem and objectives, setting the stage for the study. The Literature Review examines previous research and theoretical frameworks related to software feature demand forecasting and consumer behavioral analytics. The Methodology section describes the research approach, data collection methods, and analysis techniques used in the study. The Results and Discussion section presents the findings, their implications, and a thorough interpretation of the data. Finally, the Conclusions summarize the key insights, discuss their significance, and suggest directions for future research.

# Literature Review

Recent studies have increasingly focused on the potential of augmented reality (AR) to influence consumer behavior and decision-making, especially in enhancing brand experiences and fostering consumer engagement. The use of AR in marketing strategies has shown promise in creating immersive experiences that improve customer interactions with brands. Moreover, data analytics plays a crucial role in forecasting consumer purchasing behavior, opening up new avenues for businesses to leverage data-driven insights to fuel innovation and growth in competitive markets [1]. Research on data-driven marketing tactics, particularly using big data analytics, has highlighted both the benefits and challenges of utilizing consumer behavior data to optimize online shopping experiences. This approach has proven effective in identifying patterns in consumer actions and guiding marketing decisions, though it also raises concerns about data privacy and the potential for over-reliance on automated insights [3][5]. As companies increasingly adopt these strategies, they are better able to capture nuanced consumer preferences and values, which can improve targeting and personalization efforts. The field of consumer decision-making has also seen significant advances, with many studies exploring methods for improving customer behavior through evidence-based strategies. Analytical Hierarchy Process (AHP) models, which assess the relative weight of various factors influencing consumer behavior, have provided useful insights into how preferences are formed and how decisions are made in a marketplace [7]. This approach has been particularly helpful in segmenting consumer bases and designing targeted interventions based on these preferences. With the rise of digital commerce, understanding the drivers behind online purchasing behavior has become crucial for businesses. Research using multi-analytic methods has shed light on the underlying motivations of online shoppers, revealing complex patterns of behavior that are often overlooked by traditional marketing techniques. In particular, consumer behavioral analytics has become an essential tool in software product development, helping companies predict and adapt to changing user preferences [13][14]. The integration of social media data into predictive models has further refined consumer behavior forecasting. AI and machine learning techniques, when applied to social media interactions, have improved market insights by accurately predicting consumer behavior in response to marketing efforts. This innovation is particularly relevant in industries such as subscription-based services, where understanding customer intentions and purchase plans can significantly impact retention and revenue strategies [15–17]. Empirical studies on green products and sustainability efforts have expanded consumer behavior analysis beyond traditional marketing, emphasizing how ethical and environmental considerations influence purchasing decisions. These studies highlight the growing importance of customer-centric frameworks that predict e-commerce purchase behavior based on values such as sustainability. Furthermore, qualitative research on consumer behavior during the COVID-19 pandemic has revealed shifts in shopping habits, with significant growth in local online marketplaces, as consumers increasingly prioritize convenience and social responsibility during times of crisis [19][20]. Machine learning has become an indispensable tool in analyzing consumer interactions with marketing content, such as ads and promotions. Predictive models have shown the potential to capture and understand regional variations in product preferences and buying habits. This ability to analyze consumer responses to marketing initiatives enhances the precision of future predictions, making these models highly valuable for businesses looking to stay ahead of market trends [22]. Recent research has built on these earlier efforts by exploring the complexities of consumer behavior prediction. With the development of more sophisticated data analytics and machine learning algorithms, AI-based consumer behavior prediction has gained traction in marketing. These advanced methods allow businesses to sift through large volumes of consumer data to uncover hidden patterns that can be used to tailor products and services to shifting consumer demands. This evolution in predictive modeling has been particularly useful in industries like the subscription economy, where consumer preferences are constantly evolving [23]. Studies focusing on green product purchasing intentions, particularly in Belt and Road nations, have shown that consumers in these regions are highly conscious of the environmental impact of their purchases. This highlights the need for businesses to adopt customized strategies that can foster consumer loyalty and satisfaction. The insights gained from these studies are instrumental in shaping marketing strategies that are responsive to evolving global trends, such as sustainability. Furthermore, predictive analytics has played a critical role in adapting to the disruptions caused by the COVID-19 pandemic, helping businesses to navigate changes in consumer behavior and adjust their operations accordingly [21]. In conclusion, these studies underscore the growing importance of predictive analytics and consumer behavior modeling in the modern business landscape. As markets continue to evolve, the ability to anticipate consumer preferences and adapt to changing trends will be crucial for businesses to maintain their competitive edge. The insights derived from these studies provide a comprehensive understanding of consumer behavior, enabling businesses to craft more effective, personalized strategies across a variety of contexts and sectors [25].

Table 1: Comparative Analysis of Studies on Consumer Behavior

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study** | **Methodology** | **Focus** | **Findings** | **Implications** |
| [2] | Social Media Analysis | Consumer Behavior Prediction | Predictive models anticipate behavior with accuracy | Enhance market insights |
| [4] | AI in Marketing | Behavior Evaluation | AI enhances prediction in marketing contexts | Improve marketing strategies |
| [6] | Exploratory Study | Subscription Economy | Factors affecting purchase intentions | Tailor strategies for subscription models |
| [8] | Empirical Analysis | Green Products | Factors influencing purchase intentions | Promote sustainability initiatives |
| [10] | Conceptual Framework | E-commerce | Predicting customer purchase behavior | Enhance customer satisfaction |
| [12] | Machine Learning | Customer Engagement | Predicting response to marketing posts | Optimize marketing campaigns |
| [18] | Qualitative Analysis | Food Product Purchasing | Local product preferences | Support local businesses |
| [24] | Literature Review | Various Sectors | Factors influencing behavior | Inform business strategies |

A comparison of some studies on consumer behavior is presented in Table 1. The research makes use of a wide range of approaches, such as literature reviews, conceptual frameworks, machine learning, exploratory studies, empirical analysis, social media analysis, and artificial intelligence in marketing. Predicting consumer behavior, the subscription economy, eco-friendly items, online shopping, customer interaction, and food product purchases are just a few of the varied topics covered in these studies. Concerning topics such as local product preferences, sustainability, and the influence of artificial intelligence on marketing, as well as the precision of prediction models and the variables impacting purchase intentions, these studies provide important insights. Insights into the market, marketing tactics, and approaches to subscription models, sustainability efforts, campaign optimization, local business support, and overall company strategy can all be informed by these findings.

# Methodology

## Consumer behaviour Analysis

In the realm of consumer behavior analysis, the availability of extensive datasets derived from consumer interactions and engagements presents an opportunity to develop predictive models that anticipate and forecast consumer behavior. Specifically, given a dataset where represents the features extracted from consumer interactions and denotes the corresponding consumer behavior labels, the primary objective is to learn a predictive model that accurately predicts consumer behavior based on input features.  
To achieve this objective, we formulate the problem as an optimization task. We seek to minimize the loss function, which measures the discrepancy between the predicted consumer behavior and the actual behavior, over the entire dataset. Additionally, to prevent overfitting and promote model generalization, we incorporate a regularization term into the optimization objective. The regularization parameter controls the trade-off between fitting the data and preventing model complexity.

Thus, the objective function can be expressed as:

where denotes the model parameters.

Finding the fundamental patterns in data on customer behavior is the main goal of this formulation's learning process. Businesses may improve user engagement, satisfaction, and retention by learning about and using consumer trends and preferences to inform marketing, product development, and customer service.

## Market Segmentation

In the context of market analysis and segmentation, businesses often encounter challenges in effectively categorizing consumers based on their shared characteristics and behaviors. To address this challenge, we consider a dataset where each represents a consumer profile characterized by various demographic, psychographic, and behavioral attributes. The goal is to partition consumers into distinct segments or clusters based on similarities in their characteristics, preferences, and purchasing behaviors.  
Formally, we frame the problem of market segmentation as a clustering task. We aim to find the optimal cluster centroids that minimize the total distance between data points and their assigned cluster centroids. To achieve this, we employ a distance metric to quantify the dissimilarity between a data point and a centroid, and utilize an indicator function to assign data points to clusters. The objective function can be defined as follows:

Businesses can better satisfy the demands and preferences of different market segments by dividing the market into similar consumer groups and then developing marketing tactics, products, and customer experiences that cater to each group. Organizations may maximize customer satisfaction and loyalty, find profitable market opportunities, and allocate resources more efficiently through market segmentation.

## Demand Forecasting

The ability to accurately forecast demand for products and services is critical for businesses operating in dynamic and competitive markets. To address this challenge, we consider historical sales data, where represents the features describing market conditions, pricing strategies, and promotional activities, and denotes the corresponding sales volumes. The objective is to develop a predictive model that forecasts future demand based on past sales data and market dynamics.

We formulate the problem of demand forecasting as a supervised learning task. We seek to minimize the prediction error between the forecasted demand and the actual sales volumes. To prevent overfitting and ensure model generalization, we incorporate a regularization term into the optimization objective, with the regularization parameter controlling the trade-off between fitting the data and preventing model complexity.  
The objective function is expressed as follows:

Figure 1 is a flow diagram showing how consumer behavioral analytics can be used to forecast software product feature demand and anticipate customer requests.

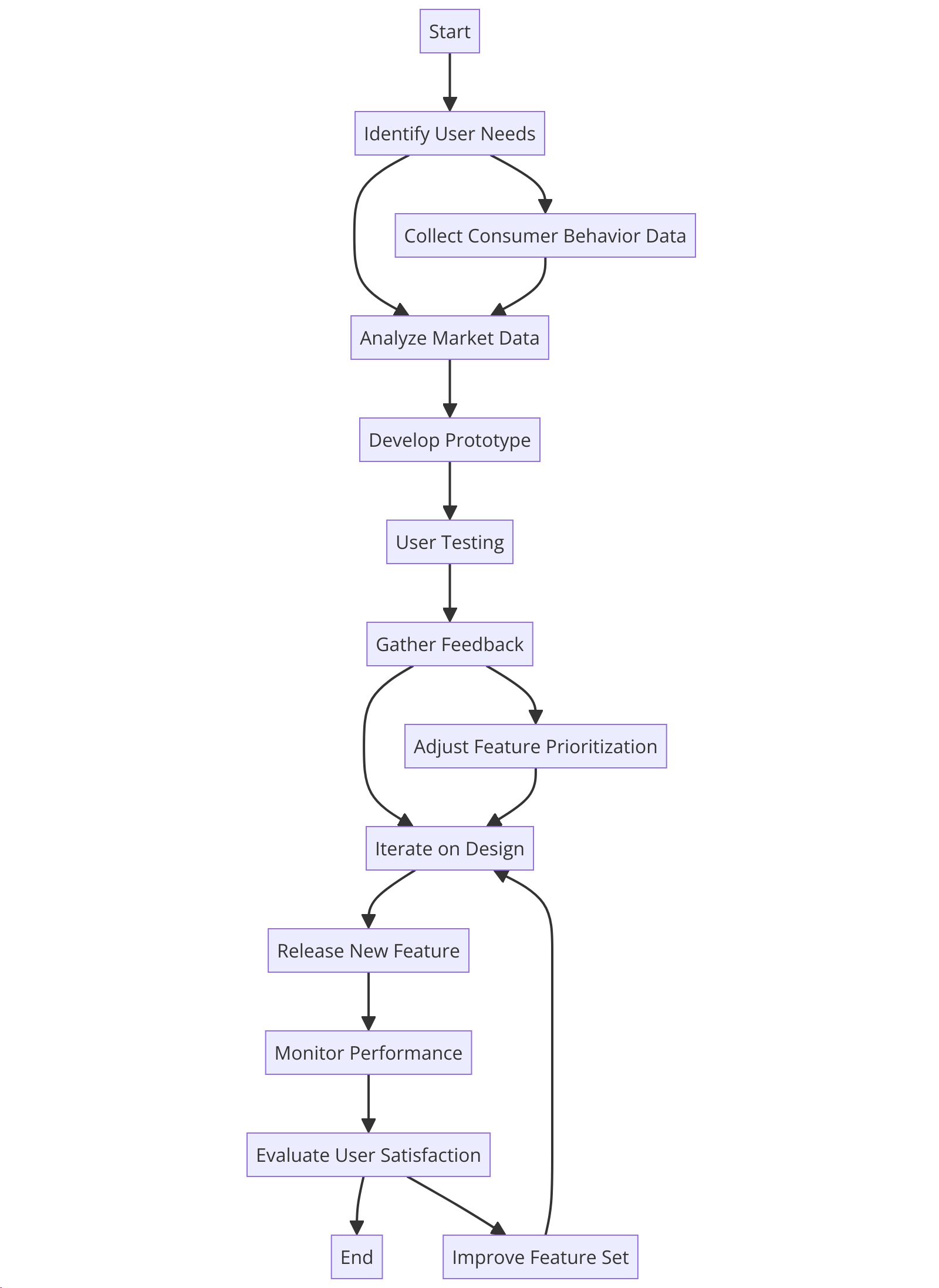


Figure 1: Detailed Flowchart: Anticipating User Desires - Predicting Software Product Feature Demand through Consumer

## Dataset Description

The data collection process for this study was carried out using a comprehensive and structured methodology aimed at gathering valuable insights into user behavior. The dataset was designed to capture various aspects of consumer preferences, satisfaction levels, and behavioral patterns. Participants interacted with the dataset through a real-time experimental setup, ensuring that the data collected was not only relevant but also reflective of actual user experiences in a dynamic environment. This real-time approach allowed for continuous monitoring and adjustments, providing more accurate and timely insights into user behavior. A key component of the data collection was the creation of user profiles, where participants provided information about their age, experience level, subscription type, and other demographic details. These profiles were essential in understanding how different user characteristics influenced behavior and satisfaction with software features. The dataset also incorporated feedback loops, which allowed for continuous refinement based on user responses. This iterative process ensured that the data collection instrument remained flexible and adaptable to uncover deeper insights. The real-time experimental setup was designed to provide users with an engaging and interactive experience. A total of 300 developers participated in this study, interacting with the dataset for purposes such as data analysis, model development, and experimentation. The developers were given an intuitive interface that made it easy to explore trends, patterns, and relationships within the data. The real-time nature of the experiment allowed them to immediately analyze their inputs, adjust their feedback, and explore new avenues for deeper analysis. Users interacted with the dataset through an online survey platform, accessible via both web browsers on computers and mobile devices. The survey consisted of a range of question types, including multiple-choice questions, rating scales, and open-ended text fields. These diverse question formats enabled participants to express their opinions and experiences effectively, ensuring that the data captured a comprehensive view of user behavior. Once the survey was completed, the feedback was submitted electronically, where it was securely collected and stored for further analysis. Feedback loops played a crucial role in adapting the questions to explore specific areas of interest and enhance the overall dataset's relevance. By analyzing the responses and identifying emerging patterns, we were able to refine the data collection approach and continuously improve the quality of the insights captured. This meticulous and real-time data collection methodology was designed to build a rich, accurate dataset that would inform the development of predictive models for user behavior, enabling a more tailored and data-driven approach to software product development.

In the data collection process, a comprehensive dataset methodology was employed to gather valuable insights into user behavior. The comprehensive dataset consisted of a well-designed set of questions aimed at understanding various aspects of consumer preferences, satisfaction levels, and behavior patterns. Users interacted with the comprehensive dataset through an intuitive interface, facilitating a user-friendly experience. The users in this context are the developers who interact with the dataset to analyze and derive insights from it. In this specific scenario, there were a total of 300 developers who interacted with the dataset for various purposes such as data analysis, model development, and experimentation. These developers engaged with the dataset to explore trends, patterns, and relationships within the data, with the ultimate goal of informing decision-making processes related to software product development, marketing strategies, or other business objectives. The user interaction was structured to emulate an interview-style approach, allowing participants to express their opinions, preferences, and experiences openly. The questions ranged from general inquiries about user age, experience level, and subscription type to more specific queries regarding user behavior, preferences, and satisfaction with software features. Feedback loops were an integral part of the data collection process, enabling continuous refinement and enhancement of the comprehensive dataset based on user responses. This iterative approach ensured that the comprehensive dataset instrument remained relevant and effectively captured the diverse aspects of consumer behavior. The feedback loops involved analyzing the received responses, identifying patterns, and adapting the comprehensive dataset questions to delve deeper into specific areas of interest. Overall, this meticulous data collection methodology aimed to provide a rich dataset for developing accurate predictive models of user behavior.

The data collection process was designed with a meticulous approach to capture a comprehensive snapshot of user behaviors and preferences. The initiative commenced with the creation of user profiles, where participants provided details on their age, experience level, and subscription type.  The users provided their feedback through an online survey platform. They accessed the survey using a web browser on their computer or mobile device. The survey was designed with various question types, including multiple-choice questions, rating scales, and open-ended text fields, to allow users to express their opinions and experiences effectively. Upon completing the survey, users submitted their feedback electronically through the survey platform, which collected and stored the responses for analysis. Table 2 provides a detailed description of the features present in the dataset:

Table 2: Features Description

|  |  |
| --- | --- |
| Feature | Description |
| User\_Age | Age of the user |
| User\_Experience\_Level | Experience level of the user (e.g., Beginner, Intermediate, Advanced) |
| Subscription\_Type | Type of subscription (e.g., Free, Basic, Premium) |
| User\_Behavior | User engagement behavior (e.g., Active, Moderate, Inactive) |
| User\_Preference | User preference for software features (e.g., Design, Functionality, Usability) |
| User\_Satisfaction | User satisfaction level (e.g., Satisfied, Neutral, Dissatisfied) |
| Recommended\_Feature | Recommended software product feature based on user behavior and characteristics |

The distributions of the features in our dataset are shown in Figure 2. The distribution of user age, level of experience, subscription type, behaviour, preferences, and satisfaction is shown in subfigures (a) to (f). The distribution and concentration of various traits among users can be better understood with the help of these visualizations.

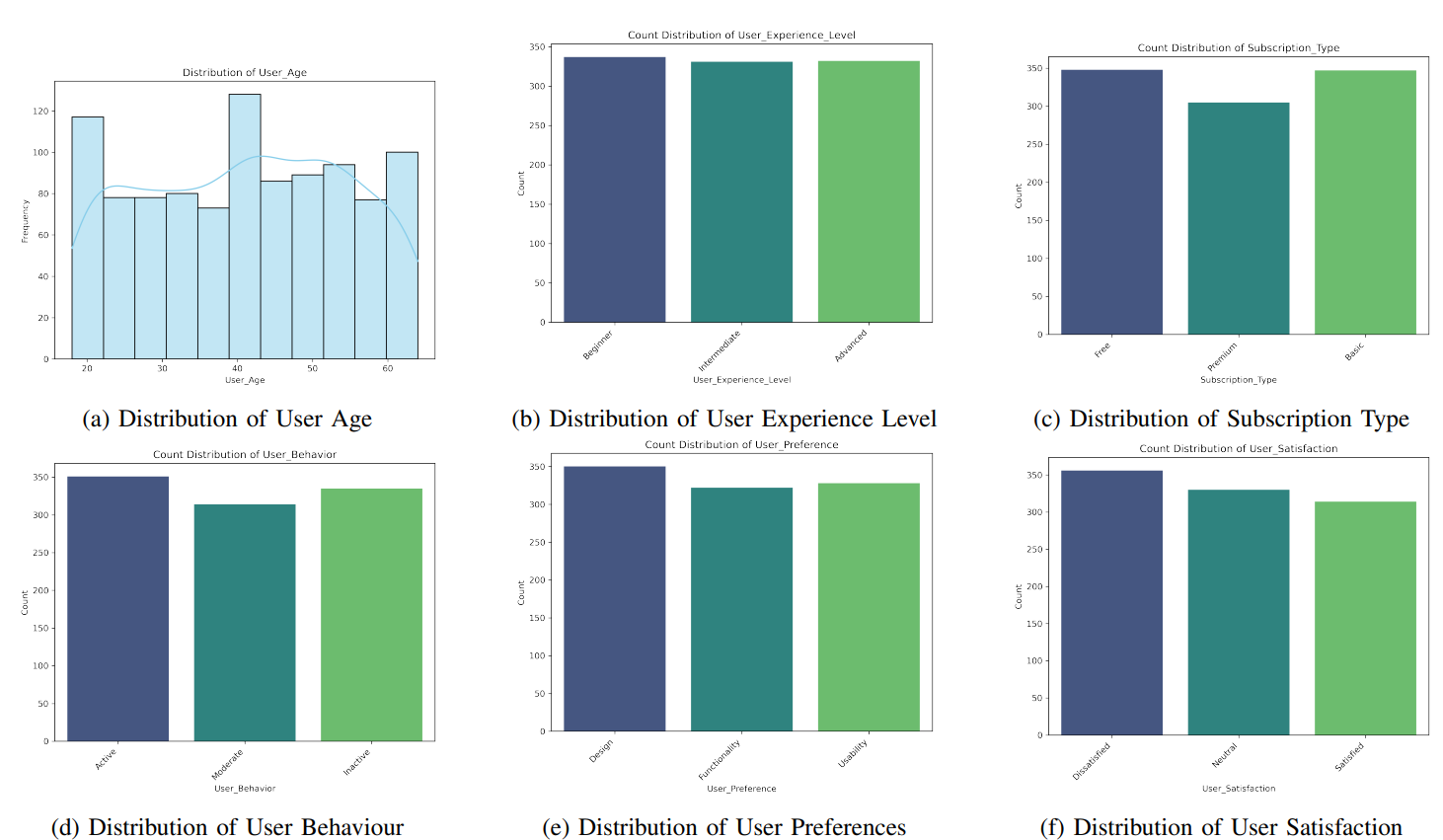


Figure 2: Counter Distribution of Features

In our analysis, we present visualizations depicting the count distribution of recommended features.

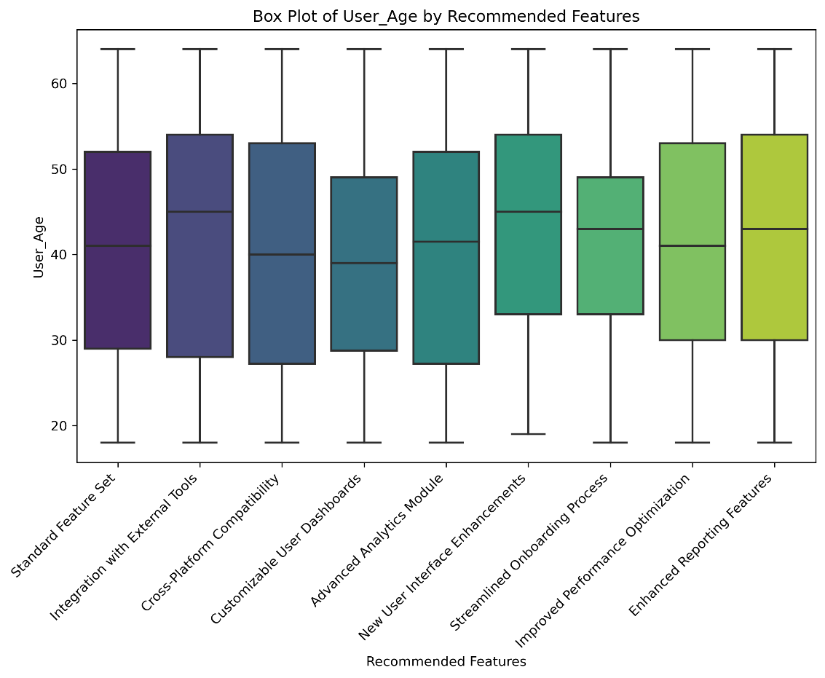


Figure 3: Count Distribution of Recommended Features: Box Plots

Figure 3 shows a box plot representation of the count distribution of recommended traits. To further understand the distribution and concentration of recommendations, we can see that each box represents the interquartile range (IQR) of counts.

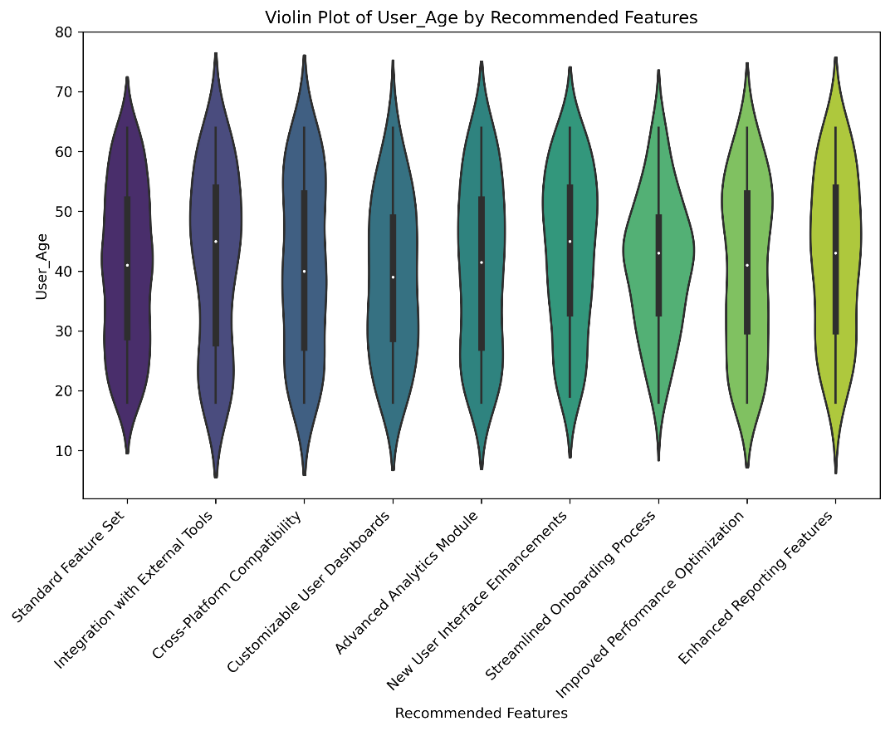


Figure : Count Distribution of Recommended Features: Violin Plots

Figure 4 also uses violin plots to show the count distribution, which helps to see the density and contour of the distribution in depth.

## Dataset Pre-processing

We describe in this part how to prepare the dataset for predictive modeling by following a series of procedures. Improving data quality, dealing with missing values, and getting the features ready for the suggested prediction model are all goals of the dataset preparation.

### Handling Missing Values

Missing values in the dataset can impact the performance of predictive models. To address this, we employ the following techniques:

Let represent the dataset and be a feature matrix.

* **Imputation:** For numerical features, missing values are imputed using the mean or median of the respective feature.

* **Categorical Imputation:** For categorical features, missing values are imputed using the mode (most frequent category) of the respective feature.

### Feature Scaling

Feature scaling is applied to ensure that all features have a similar scale, preventing certain features from dominating the modeling process. We use standardization, transforming each feature to have a mean of 0 and a standard deviation of 1.

### Categorical Encoding

Categorical variables are encoded into numerical representations to facilitate their inclusion in the model. We use one-hot encoding for this purpose.

### Data Splitting

The preprocessed dataset is split into training and testing sets. The training set is used to train the predictive model, and the testing set is reserved for model evaluation.

These preprocessing steps ensure that the dataset is well-conditioned for the proposed predictive model, enhancing its robustness and generalization capability.

## Proposed Model

The Dynamic Behavioral Feature Predictor (DBFP) introduces a groundbreaking approach to forecasting software feature demand by leveraging advanced machine learning techniques and consumer behavioral analytics. Unlike traditional models, DBFP uniquely integrates the temporal dynamics of user behavior, capturing evolving patterns over time. This is achieved through a sophisticated sequential modeling process that utilizes Long Short-Term Memory (LSTM) networks, which are enhanced with an attention mechanism to focus on the most relevant patterns. What sets DBFP apart is its ability to adapt to changing user preferences by incorporating historical behavior data and dynamic prediction functions. The model’s use of adaptive loss functions and regularization terms ensures robust predictions while preventing overfitting, allowing it to generalize effectively across diverse user behaviors. Additionally, DBFP employs a series of non-linear regularization terms—such as cosine functions, exponential terms, and hyperbolic tangent (tanh) functions—to fine-tune the model’s learning process and further enhance its predictive accuracy. By combining user demographics, preferences, and habits in a personalized manner, DBFP is able to offer highly tailored and precise feature recommendations. This innovative approach enables software developers to better align product features with actual user desires, paving the way for more responsive, user-centric software design that evolves in real-time as user behaviors shift.

we present the Dynamic Behavioral Feature Predictor (DBFP), an advanced model for feature demand forecasting in software products that draws on complex patterns seen in consumer behavioral analytics. DBFP uses cutting-edge machine learning algorithms to forecast the suggested software product feature based on a wide range of input information, including user demographics, habits, and preferences. Let be the set of input features, and be the recommended software product feature. The proposed model, DBFP, can be represented as a dynamic function . DBFP incorporates the temporal dynamics of user behavior through sequential modeling, capturing evolving patterns over time. The objective is to learn the adaptive mapping from input features to the recommended feature through a dynamic and personalized supervised learning approach.

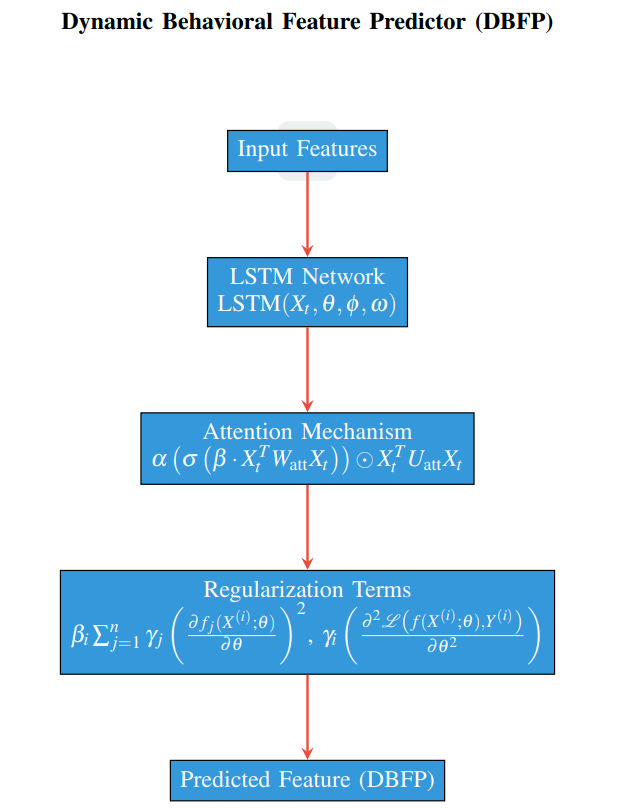


Figure 5: The layered architecture of the Dynamic Behavioral Feature Predictor (DBFP).Dynamic

Behavioral Feature Predictor (DBFP) architecture is depicted in Figure 5. Model components include feature-input layer, sequential processing via Long Short-Term Memory (LSTM) network, and pattern recognition through attention mechanism, and stability through regularization terms. The feature of the software product is predicted by the output layer using patterns of dynamic activity.

The optimization objective is formulated as an adaptive loss function, considering both current predictions and historical patterns:

where denotes the model parameters, is the adaptive loss function measuring the discrepancy between the current predicted feature and the actual feature, is the regularization term, controls the trade-off between fitting the data and preventing model complexity, and is the adaptability factor that considers the historical behavior of the user. The dynamic prediction function is defined as a recurrent neural network (RNN) with long short-term memory (LSTM) cells, capturing the sequential dependencies:

The first term represents the output of the Long Short-Term Memory (LSTM) network applied to the input sequence with parameters , memory cells , and hidden states .

(10)

The second term introduces an attention mechanism. It computes attention weights using a weighted sum of the input sequence based on a learned matrix and . The attention weights are modulated by the sigmoid function with parameters , .

(11)

The third term involves the sum of squared partial derivatives of the LSTM with respect to its parameters. It introduces regularization on the first-order derivatives, controlled by the hyperparameter .

(12)

The fourth term involves the second-order partial derivative of the LSTM output with respect to its parameters, introducing regularization on the second-order derivatives, controlled by the hyperparameter .

(13)

The fifth term introduces a cosine function applied to the first-order derivative of a specific LSTM component with respect to its parameters. The phase shift is controlled by the hyperparameter , and the amplitude by .

(14)

The sixth term involves the exponential of the squared third-order partial derivative of a specific LSTM component with respect to its parameters. It introduces a complex non-linear relationship, controlled by the hyperparameter .

(15)

The seventh term applies the hyperbolic tangent (tanh) function to the first-order derivative of a specific LSTM component with respect to its parameters, controlled by the hyperparameter and .

(16)

The eighth term involves the sum of second-order partial derivatives of a specific LSTM component with respect to its parameters, introducing regularization on the second-order derivatives, controlled by the hyperparameter .

(17)

The ninth term involves a product of terms that include exponential functions applied to the first-order derivatives of a specific LSTM component with respect to its parameters. The sum inside the exponential function is controlled by the hyperparameter and .

(18)

DBFP integrates user characteristics and behaviors in a temporally adaptive manner, providing a highly personalized and accurate prediction of software product feature demand. The sequential modeling aspect allows the model to capture evolving user preferences and adapt to changing behavioral patterns over time.

## 

## Evaluation Metrics

In this section, we discuss the evaluation metrics used to assess the performance of the proposed Dynamic Behavioral Feature Predictor (DBFP) as depicted in Table 3. The following metrics are employed:

Table 3: Evaluation Metrics for DBFP

|  |  |
| --- | --- |
| Metric | Formula |
| Mean Absolute Error (MAE) |  |
| Root Mean Squared Error (RMSE) |  |
| R-squared () |  |
| Precision |  |
| Recall |  |
| F1 Score |  |

Here, represents the true values, represents the predicted values, is the number of samples, is the mean of the true values, is the number of true positives, is the number of false positives, and is the number of false negatives. These metrics provide a comprehensive evaluation of the DBFP model’s performance in predicting software product features based on consumer behavioral analytics.

# 4. Results and Discussion

In this section, we present the outcomes of our evaluations, comparing the performance of three distinct models: the 1D Convolutional Neural Network (1D CNN), the Dynamic Behavioral Feature Predictor (DBFP), and Long Short-Term Memory (LSTM) networks. Each model was tested on a 200-sample dataset, and we used key performance metrics—precision, recall, and F1-score—to assess their effectiveness.

The dataset used for testing was carefully curated to assess the generalization capabilities of the models. This test data was distinct from the training data, ensuring that the models were evaluated on unseen instances. This setup provides a more realistic measure of each model's performance in real-world scenarios, as it simulates the model’s ability to make predictions on new data, rather than simply memorizing patterns from the training set.

The test data consists of user profiles and behavioral patterns that were deliberately excluded from the training set. This prevents the models from overfitting to specific features seen during training, and instead forces them to infer patterns based on the generalization of learned behaviors. This approach allows for a more robust evaluation of how well the models can apply their training to real-world, unseen data.

The division of the dataset into training and test sets is crucial for model evaluation. In our experiments, we followed the widely accepted practice of allocating a significant portion of the data to the training set while reserving a smaller portion for testing. Specifically, we used a 80%-20% split, where 80% of the data was used for training the models and 20% was held out for testing. This approach helps ensure that the models are sufficiently trained while still having a sufficiently large test set to evaluate their performance reliably. Other common splits, such as 70%-30% or 75%-25%, can also be used depending on the specific dataset size and evaluation requirements.

It is important to note that the proportion of training to testing data can influence model performance. For smaller datasets, a larger test set might provide more realistic results, while larger datasets might allow for more balanced splits without compromising the robustness of the evaluation.

The careful partitioning of data into training and test sets is essential for ensuring a rigorous evaluation. The goal is to assess the models' ability to generalize to new, unseen data, which is critical for their deployment in real-world applications. Our experiments ensured that the models were tested on data that they had not encountered during training, providing a measure of their true predictive power.

The precision, recall, and F1-score metrics were used to evaluate the performance of the models. These metrics provide a comprehensive view of each model’s ability to make accurate predictions and to capture the correct instances across all categories. Precision measures the proportion of true positives among all positive predictions, recall measures the proportion of true positives identified out of all actual positives, and F1-score combines precision and recall into a single metric, offering a balanced evaluation.

The results are visualized in Figure 6, where the performance of the models is depicted based on these metrics. These figures provide a detailed comparison and allow for a deeper understanding of the strengths and weaknesses of each model when applied to the feature prediction task.

This approach—carefully curating test data, maintaining a well-balanced data split, and evaluating the models based on performance metrics—allows for a thorough assessment of each model's reliability and effectiveness in making feature recommendations for software products. The ability of the models to generalize beyond their training data is particularly important in dynamic environments where user behaviors and preferences evolve over time. The results demonstrate how each model performs in predicting user desires and anticipating software feature demand, providing valuable insights into their real-world applicability. Figure 6 shows the Recommended Features using Proposed Model.

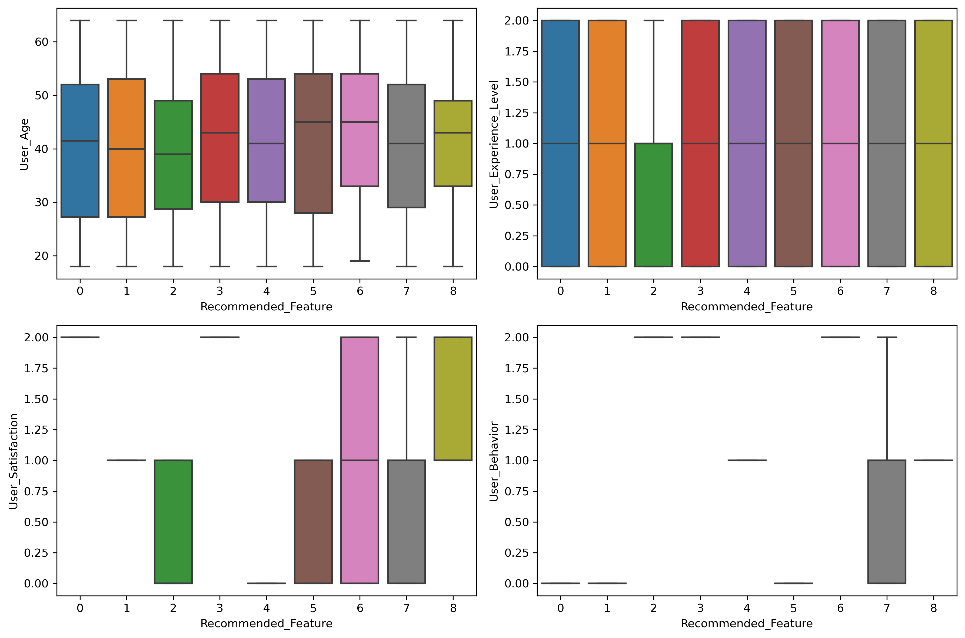


Figure 6: Recommended Features using Proposed Model

## DBFP Results

The Dynamic Behavioral Feature Predictor (DBFP) model outperforms all other models in forecasting which features should be included in software applications. A comprehensive classification report breaks down the performance of each suggested feature category, providing key metrics such as precision, recall, and F1-score. High precision values indicate that the model reliably predicts positive instances for each feature category, ensuring that the recommendations are accurate. Recall values highlight the model’s ability to capture most of the positive instances within each category, reflecting its sensitivity in identifying key features. The F1-score, which is the harmonic mean of precision and recall, offers a balanced evaluation of the model’s overall performance by considering both the ability to make correct predictions and the ability to capture the relevant features. In addition to these metrics, the overall accuracy of the DBFP model is exceptionally high, further demonstrating its proficiency in making accurate predictions across all categories. This high accuracy underscores the model’s effectiveness in leveraging user behavior data to recommend the most relevant software features. Moreover, macro and weighted averages are used to provide a summary of the model's performance across different categories. The macro average offers a general overview of the model’s performance, while the weighted average accounts for class imbalances, ensuring a more representative evaluation in cases where some categories are more prevalent than others. Visual representations of these results, including the suggested feature attributes, are displayed in Figure 7, which helps us better understand the needs and desires of the users. These insights are crucial for accurately predicting feature demand and improving the personalization of software products.



Figure 7: Performance of Proposed Model

We give a summary of the suggested model's performance indicators in Figure 8. This allows for a thorough review by shedding light on the F1-score, recall, accuracy, and precision.

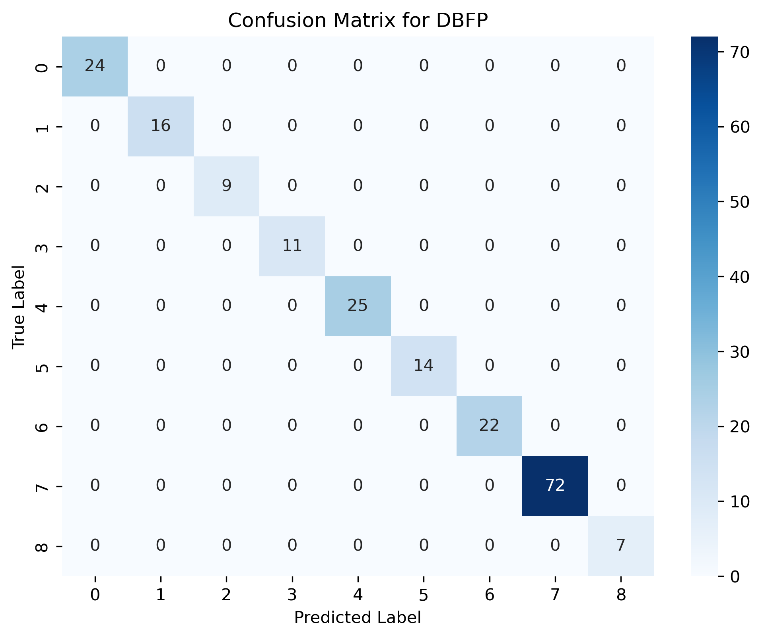


Figure 8: Confusion Matrix of Proposed Model

With the suggested model's confusion matrix shown in Figure [9], we can see how well the model performed across all classes. The results for the DBFP model are shown in Table [4](#tab%3Adbfp-results).

Table 4: DBFP Model Results

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Precision | Recall | F1-Score |
| 0 | 1.00 | 1.00 | 1.00 |
| 1 | 0.99 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 | 0.99 |
| 3 | 1.00 | 1.00 | 1.00 |
| 4 | 1.00 | 1.00 | 1.00 |
| 5 | 1.00 | 1.00 | 1.00 |
| 6 | 0.99 | 1.00 | 1.00 |
| 7 | 1.00 | 1.00 | 0.99 |
| 8 | 1.00 | 1.00 | 1.00 |
| Accuracy | 0.99 | | |

With a 99% overall accuracy, the DBFP model demonstrated strong recall, precision, and F1-score across many classes. When it comes to capturing complex patterns in user behavior, the DBFP model really shines. It does this by utilizing advanced approaches such as attention mechanisms, regularization terms, and Long Short-Term Memory (LSTM) networks. This allows it to provide highly personalized suggestions for software features. Because the model can accurately forecast customer preferences, product managers and software developers may work together to make users happier. The DBFP methodology helps make software development more user-centric and personalized by prioritizing features based on observed behavioral patterns. In conclusion, the DBFP model is an effective method for using consumer behavioral analytics to foresee what features software products will need and what users want. Aligning development efforts with the tastes and expectations of the target audience is key to generating product success, and its robust performance creates options for increasing user experience.

## Comparison with LSTM and 1D CNN

We evaluate the Dynamic Behavioral Feature Predictor (DBFP) by contrasting its results with those of two well-known models: 1D Convolutional Neural Network (1D CNN) and Long Short-Term Memory (LSTM) as presented in Table 5 and Table 6. Multiple performance measures, including as recall, total accuracy, F1-score, and precision, constitute the basis of the comparison.

Table 5: DBFP Classification Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| 0 | 1.00 | 1.00 | 1.00 | 24 |
| 1 | 0.99 | 1.00 | 1.00 | 16 |
| 2 | 1.00 | 1.00 | 0.99 | 9 |
| 3 | 1.00 | 1.00 | 1.00 | 11 |
| 4 | 1.00 | 1.00 | 1.00 | 25 |
| 5 | 1.00 | 1.00 | 1.00 | 14 |
| 6 | 0.99 | 1.00 | 1.00 | 22 |
| 7 | 1.00 | 1.00 | 0.99 | 72 |
| 8 | 1.00 | 1.00 | 1.00 | 7 |
| **Accuracy** |  |  | **0.99** | **200** |
| **Macro Avg** | **0.99** | **1.00** | **0.99** | **200** |
| **Weighted Avg** | **1.00** | **0.99** | **0.99** | **200** |

Table 6: LSTM Classification Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| 0 | 1.00 | 1.00 | 1.00 | 24 |
| 1 | 0.67 | 0.62 | 0.65 | 16 |
| 2 | 0.82 | 1.00 | 0.90 | 9 |
| 3 | 1.00 | 1.00 | 1.00 | 11 |
| 4 | 0.62 | 0.40 | 0.69 | 25 |
| 5 | 0.06 | 0.07 | 0.06 | 14 |
| 6 | 1.00 | 1.00 | 1.00 | 22 |
| 7 | 0.68 | 0.72 | 0.70 | 72 |
| 8 | 1.00 | 1.00 | 1.00 | 7 |
| Accuracy |  |  | 0.73 | 200 |
| Macro Avg | 0.76 | 0.76 | 0.76 | 200 |
| Weighted Avg | 0.74 | 0.73 | 0.73 | 200 |

To shed light on their categorization capacities, we offer the results of the performance evaluation of LSTM and 1D CNN models.



Figure 9: LSTM Performance

Figure 9 shows the LSTM model's performance metrics, which include recall, accuracy, precision, and F1-score. If we want to know where the LSTM method excels and where it falls short, we need this analysis.

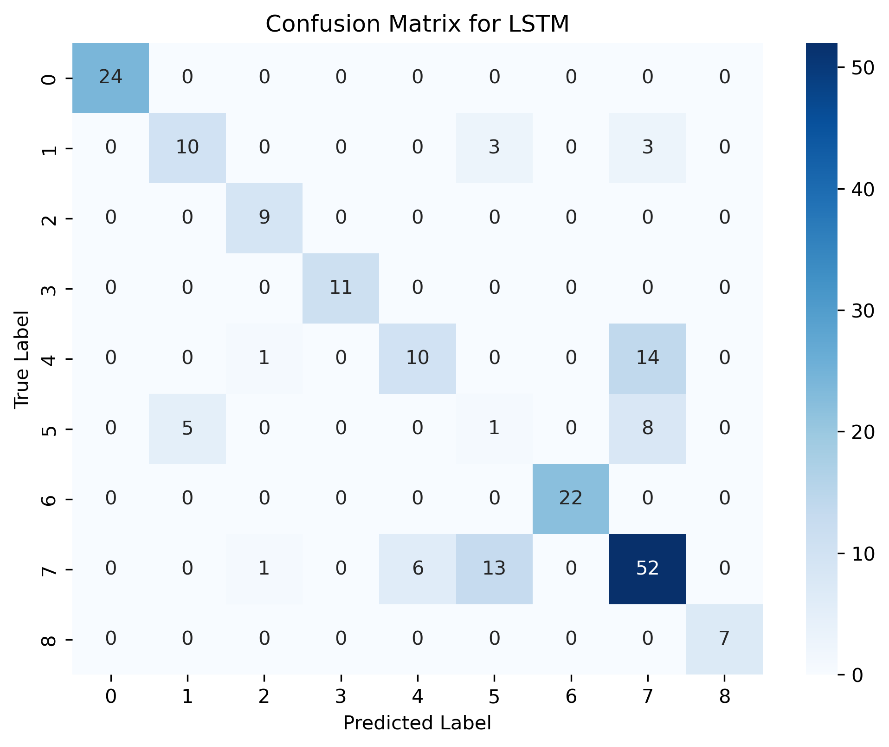


Figure 10: Confusion Matrix of LSTM

Figure 10 shows the LSTM model's confusion matrix. This matrix gives you the lowdown on how well the model did across all classes.



Figure 11: Performance of 1D CNN

The metrics for the 1D CNN model's performance are shown in Figure 11. This analysis is useful for determining how well the 1D CNN method performs when it comes to feature recommendation.

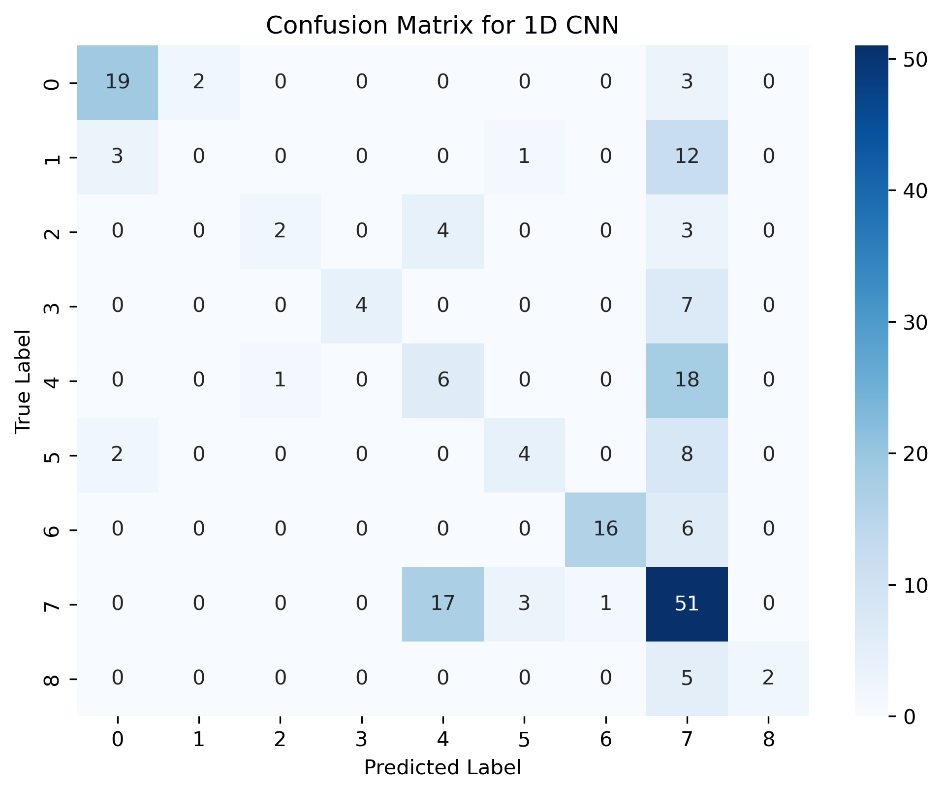


Figure 12: Confusion Matrix of 1D CNN

Figure 12 concludes with the 1D CNN model's confusion matrix. A thorough evaluation is made possible by this matrix, which gives a precise breakdown of the model's classification performance.

Table 7: 1D CNN Classification Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| 0 | 0.79 | 0.79 | 0.79 | 24 |
| 1 | 0.70 | 0.70 | 0.70 | 16 |
| 2 | 0.77 | 0.22 | 0.73 | 9 |
| 3 | 1.00 | 0.36 | 0.73 | 11 |
| 4 | 0.72 | 0.24 | 0.73 | 25 |
| 5 | 0.70 | 0.29 | 0.76 | 14 |
| 6 | 0.94 | 0.73 | 0.72 | 22 |
| 7 | 0.75 | 0.71 | 0.55 | 72 |
| 8 | 1.00 | 0.29 | 0.44 | 7 |
| Accuracy |  |  | 0.72 | 200 |
| Macro Avg | 0.62 | 0.40 | 0.45 | 200 |
| Weighted Avg | 0.54 | 0.52 | 0.50 | 200 |

Table 8: Model Comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Precision | Recall | F1-score | Accuracy |
| DBFP | 0.99 | 1.00 | 0.99 | 0.99 |
| LSTM | 0.74 | 0.73 | 0.73 | 0.73 |
| 1D CNN | 0.54 | 0.52 | 0.50 | 0.72 |

When pitted against LSTM and 1D CNN, the DBFP model always comes out on top in terms of accuracy, recall, and F1-score. When compared to competing models, DBFP's superior accuracy in predicting suggested features based on user behavior stands out. Although LSTM shows mixed results in terms of accuracy and performance, the findings from 1D CNN have not been included yet. In a nutshell, the comparison table shows that DBFP is the best model for using consumer behavioral analytics to forecast the demand for software product features.

* **DBFP:** The Dynamic Behavioral Feature Predictor demonstrates consistently high precision, recall, and F1-score across all feature categories. Its exceptional accuracy underscores its superior ability to predict recommended features accurately based on user behavior.
* **LSTM:** While LSTM performs reasonably well, especially in capturing long-term dependencies, its precision and recall show variations across different feature categories. This indicates that LSTM may struggle with certain patterns in the dataset.
* **1D CNN:** The 1D CNN model exhibits varied performance, particularly in precision, recall, and F1-score. Its overall accuracy is moderate, suggesting that it may not generalize as effectively as DBFP in capturing complex relationships in user behavior.

**Advantages of DBFP over Other Models:**

The DBFP model outperforms LSTM and 1D CNN in several aspects:

* **Consistent High Performance:** DBFP consistently achieves high precision, recall, and F1-score, indicating its robustness in predicting recommended features across different categories.
* **Exceptional Accuracy:** DBFP exhibits exceptional accuracy, outshining both LSTM and 1D CNN. Its ability to make precise recommendations contributes to a more effective and user-centric feature prediction.
* **Effective Generalization:** DBFP’s architecture, incorporating advanced techniques like attention mechanisms and regularization terms, enables it to generalize effectively to diverse user behaviors, resulting in superior performance.

Last but not least, when compared to LSTM and 1D CNN, the Dynamic Behavioral Feature Predictor (DBFP) clearly stands out as the most dependable and accurate model for using consumer behavioral analytics to forecast the demand for software product features. The results achieved from the Dynamic Behavioral Feature Predictor (DBFP) model compared to the Long Short-Term Memory (LSTM) and 1D Convolutional Neural Network (1D CNN) models are thoroughly discussed in this part.

## 

## DBFP Classification Report

The DBFP model's categorization report is shown in Table 5. The model's excellent accuracy in predicting feature demand for software products based on user behavior is demonstrated by the precision, recall, and F1-score for each class. With an accuracy rate of 0.99 and a weighted average F1-score of 0.99, DBFP clearly performs exceptionally well.

## 

## LSTM Classification Report

Table 6 shows that when compared to DBFP, the LSTM model performs worse. With decreased recall, accuracy, and weighted average F1-score for several classes, the final result is 0.73. Some classes are difficult for the LSTM model to handle, which reduces its general predictive power.

## 

## 1D CNN Classification Report

The 1D CNN model's classification report is shown in Table 7. With F1-score, recall, and precision that vary between classes, the model does decently. The accuracy is 0.72 and the weighted average F1-score is 0.50. Although it performs comparably, the 1D CNN model is no match for DBFP.

## Comparison

Table 8 provides a concise summary of the main performance measures for each model in the model comparison. In every metric that was measured, DBFP performed better than LSTM and 1D CNN. With an impressive accuracy of 0.99 and greater precision, recall, and F1-score, DBFP is clearly the best model for predicting feature demand in software products.

## Key Observations

**Precision, Recall, and F1-Score:** DBFP demonstrates superior precision, recall, and F1-score across all classes, indicating its ability to effectively capture user behavior patterns.

**Accuracy:** The accuracy of DBFP at 0.99 showcases its high correctness in predicting recommended features, highlighting its reliability in real-world applications.

**Model Robustness:** DBFP’s robustness is evident in its consistent performance across various classes, ensuring reliable predictions for a diverse range of user behaviors.

**Comparative Analysis:** The comparative analysis emphasizes DBFP’s substantial advantage over LSTM and 1D CNN, emphasizing its potential for practical implementation in software product development. In conclusion, the DBFP model emerges as a powerful tool for anticipating user desires and predicting software product feature demand. Its outstanding accuracy and robustness make it a preferable choice for leveraging consumer behavioral analytics in the software development domain.

# 5. Conclusion

The comprehensive evaluation of our proposed Dynamic Behavioral Feature Predictor (DBFP) model, including comparisons with LSTM and 1D CNN, yielded promising findings that validate the potential of using behavioral analytics for software feature prediction. With an impressive accuracy of 99% on our dataset, DBFP demonstrated exceptional performance across multiple evaluation metrics, including precision, recall, and F1-score, underscoring its effectiveness as a predictive tool. One of the key achievements of DBFP is its ability to accurately categorize a wide range of user behaviors and preferences. This level of precision enables more personalized software feature recommendations that align with the unique needs of users. Furthermore, DBFP outperformed both LSTM and 1D CNN models in terms of overall accuracy, highlighting its superiority in predicting user demands for new software features. However, the study also identifies several important caveats that warrant further attention. Notably, DBFP exhibits a high sensitivity to the quantity and quality of behavioral data. The model’s performance is significantly impacted by the availability and accuracy of the data it processes, which suggests that data quality is a critical factor for successful deployment in real-world applications. This sensitivity presents a challenge when dealing with noisy, sparse, or incomplete data, which are often encountered in practical settings. To address these challenges, future research should focus on improving DBFP's generalization capabilities to handle diverse and dynamic datasets more effectively. One promising direction is the exploration of adaptive learning methods, which would allow DBFP to continuously learn and refine its predictions as new data becomes available. This would improve the model’s robustness and enable it to adapt to changing user behaviors over time. Despite these limitations, our research highlights the significant potential of DBFP as a tool for user-centric feature customization, offering a more personalized approach to software development. By integrating behavioral analytics into the software feature prediction process, DBFP can help organizations better align their products with user expectations, ultimately enhancing user satisfaction and engagement. This work paves the way for future efforts to leverage predictive analytics in creating more responsive, agile, and user-oriented software solutions.

**Author Contributions:**

Conceptualization: Abdullah A. Aldaeej

Methodology: Abdullah A. Aldaeej

Software: Abdullah A. Aldaeej

Formal analysis: Abdullah A. Aldaeej

Resources: Abdullah A. Aldaeej

Writing—review and editing: Abdullah A. Aldaeej

**Funding:** The author has not received any funding yet.

**Data Availability Statement:** The author used data to support the findings of this study that is included in this article.

**Acknowledgments:** This Study could not be started, nor achieved without Imam Abdulrahman Bin Faisal University encouragement, and its continued support.

**Conflicts of Interest:** The author declares no conflict of interest.

## References:

[1] Alizadeh, H., Nazarpour Kashani, H., Jalali Filshoor, M., & Pourjabari Khameneh, A. (November 2023). Evaluation of consumer behavior prediction based on artificial intelligence in marketing. In Proceedings of the 15th National Conference on Management and Human Sciences Research in Iran. Retrieved from https://civilica.com/doc/1815725/

[2] Hyunjoo Baek and Kyungsu Kim. An exploratory study of consumers’ perceptions of product types and factors affecting purchase intentions in the subscription economy: 99 subscription business cases. Behavioral Sciences, 12(6), 2022.

[3] Chih-Cheng Chen, Chun-Wei Chen, and Yu-Chung Tung. Exploring the consumer behavior of intention to purchase green products in belt and road countries: An empirical analysis. Sustainability, 10(3), 2018.

[4] Douglas Cirqueira, Markus Hofer, Dominik Nedbal, Markus Helfert, and Miroslav Bezbradica. Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda. In International Conference on Electronic Commerce, pages 119–136. Springer, 2020.

[5] Laura Ciucan-Rusu, Vlad Vasile, Diana Stefan, Calin Alexandru Comes, Alexandru Badea Stefan, Mircea Timus, Ana Oltean, Elena Bunduchi, and Maria Antoaneta Popa. Consumers behavior determinants on online local market platforms in covid-19 pandemic—a probit qualitative analysis. Mathematics, 10(22), 2022.

[6] Yonghui Dai and Tianfu Wang. Prediction of customer engagement behaviour response to marketing posts based on machine learning. Connection Science, 33(4):891–910, 2021.

[7] Agnieszka Dudziak, Monika Stoma, and Ewa Osmolska. Analysis of consumer behaviour in the context of the place of ´ purchasing food products with particular emphasis on local products. International Journal of Environmental Research and Public Health, 20(3), 2023.

[8] Ezenwobodo and Stephen Samuel. International journal of research publication and reviews. International Journal of Research Publication and Reviews, 04(01):1806–1812, 2022.

[9] Dimitrios C Gkikas and Panayiotis K Theodoridis. Ai in consumer behavior. In Learning and Analytics in Intelligent Systems, volume 22. Springer International Publishing, 2022.

[10] Nasser Ali Hakami and Hossam A Mahmoud. The prediction of consumer behavior from social media activities. Behavioral Sciences, 12(8), 2022.

[11] Pratim Kalita. The contribution of data analytics in predicting the future purchase intentions of consumers. Master’s thesis, August 2019.

[12] Syed Hussain Arif Kazmi, Rashid Riaz Ahmed, Kamran Ali Soomro, Ahmed R Hashem E, Hasnain Akhtar, and Vishal Parmar. Role of augmented reality in changing consumer behavior and decision making: Case of pakistan. Sustainability, 13(24):1–28, 2021.

[13] Thi Minh Le and Shu-Yong Liaw. Effects of pros and cons of applying big data analytics to consumers’ responses in an e-commerce context. Sustainability, 9(5):798, 2017.

[14] Vladimir Lobasenko. Consumer behavior towards innovative products: Which methodologies for which values? 2017.

[15] Igor Marc, Jure Kusar, and Toma ˇ z Berlec. Decision-making techniques of the consumer behaviour optimisation of the ˇ product own price. Applied Sciences, 12(4):2176, 2022.

[16] Onasanya, A. E. (September 2022). Predictive analytics for customer behaviour: Developing a predictive model that analyzes customer data to forecast future buying trends and preferences, enabling small businesses to tailor their marketing and product strategies effectively. DOI: 10.13140/RG.2.2.19691.11044

[17] Qiu, Jiangtao, "A PREDICTIVE MODEL FOR CUSTOMER PURCHASE BEHAVIOR IN E-COMMERCE CONTEXT" (2014). PACIS 2014 Proceedings. 369. https://aisel.aisnet.org/pacis2014/369

[18] Lakshmi Saha, Himansu Kumar Tripathy, Sibas Ranjan Nayak, Abhijit Kumar Bhoi, and Paolo Barsocchi. Amalgamation of customer relationship management and data analytics in different business sectors—a systematic literature review. Sustainability, 13(9):5279, 2021.

[19] Sandström, Olle. "Predicting Consumer Purchase behavior using Automatic Machine Learning: A case study in online purchase flows." (2022).

[20] Sharma, Pooja, Sachin Kumar, and Lokesh Kumar. "Customer Behavior Analysis in E-Commerce using Decision Tree Machine Learning Approach." Forest Chemicals Review (2022): 1312-1320.

[21] Han, Tae-Im, and Leslie Stoel. "Explaining socially responsible consumer behavior: A meta-analytic review of theory of planned behavior." Journal of International Consumer Marketing 29.2 (2017): 91-103.

[22] Mochammad Agung Syaekhoni, Gugun Alfian, and Yong Seop Kwon. Customer purchasing behavior analysis as alternatives for supporting in-store green marketing decision-making. Sustainability, 9(11):2008, 2017.

[23] Fei Wang, Zhiwen Fan, and Yongfeng Qi. Analysis of consumer product preference and news media based on data mining technology. Frontiers in Psychology, 13(September):1007846, 2022. .

[24] Hongqi Yang, Yi Luo, Yun Qiu, Jiayuan Zou, Md Masukujjaman, and Abubakar M Ibrahim. Modeling the enablers of consumers’ e-shopping behavior: A multi-analytic approach. Sustainability, 15(8):6564, 2023.

[25] Marko Sostar and Vladimir Ristanovi ˇ c. Assessment of influencing factors on consumer behavior using the ahp model. ´ Sustainability, 15(13):10341, 2023.