International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.56472/ICCSAIML25-124
Eureka Vision Publication | ICCSAIML'25-Conference Proceeding

Original Article

Re-Engaging the Dormant: A Data-Driven Framework for Cold Start Email Campaigns in Large-Scale Platforms

Preetham Reddy Kaukuntla
Data Science Glassdoor Texas, USA

Abstract - *Bulk email campaign starts at a 'cold', as marketing teams on large scale platforms face with very great difficulties. What these campaigns do is targeting those dormant users who have become inactive or disengaged from a service or a product. The main goal of this paper is to develop a data driven framework that manages cold start email campaigns by analyzing user behavior, segmentation of audiences and utilization of predictive analytics to optimize engagement and conversion. Therefore, based on the historical data, machine learning models and A/B testing, the proposed framework personalize the content of the email, the time and frequency to re-engage the dormant users.*

Keywords - *User re engagement, cold start, dormant users, data driven, predictive analytics, email marketing, large scale platforms, generative models.*

1. Introduction

Big problems arise in the fast-paced digital marketing world when trying to reengage users who have become dormant from using a platform, particularly when speaking to a company with a large-scale platform that attracts millions of users. Email campaigns are not an innovative practice, but the cold start email campaigns (emails targeting users who have been inactive for a couple of months) requires an elaborate, data conscious approach to motivate users to get back. Obviously just open rates or click through rates are insufficient in terms of what they tell us about how complex it is to bring reengaged users back. Though, a higher and more effective approach is on analyzing user's behavior and dividing them into subsets (audiences) according to your past experience, demographic data and interaction patterns to design high productive email campaigns in order to attract again users [1].

1.1 The Need for Data-Driven Cold Start Email Campaigns

It can be addressed as a challenge to re-engage the dormant users by using predictive analytics, user segmentation and personalization. During the time since these platforms came into existence, they have access to historical data like, how has a user acted in the past i.e. purchase behavior, time since last engagement or how well a user responds to emails, and based on all of this they partition the users into specific classes to which a tailored message can be sent. Since the timed uniqueness is on hand, there is a great chance that any user would come back in active engagement, if the content and offers are personalized [1]. And first cold start phase is highly important not understanding how they do it right can help you big first, so they went in and they understand the pattern of user behavior and what they say when they reply to it.

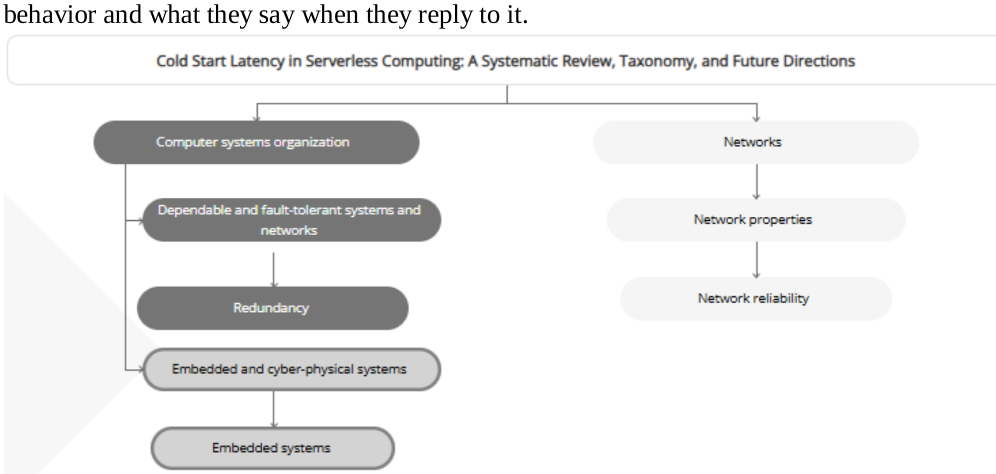

Figure 1. Cold Star Latency [2]

1.2 Leverage Machine Learning and Predictive Analytics

The aim of this research is to establish whether machine learning techniques and predictive models can be useful both in better targeting and being effective in cold start email campaigns. Machine learning algorithms can then use all of the collected data and predict which type of users will respond to some kind of email content or promotional offerings. Predictive models such as decision trees or neural networks can be used for marketing to find out what factors affect re engagement by users and to provide more targeted and personalization. Additionally, email content can be A/B tested within subject lines and call to action strategy to ensure that every email attempting to reengage dormant users is as effective as it possibly can be [2].

1.3 Purpose and Scope of Study

The direct contribution of this paper is to design a data driven framework for cold start email campaign which optimizes user reengagement through using advanced analytics, segmentation and machine learning techniques. This is in scope if it is the development of a practical, actionable strategy to allow large scale platforms to road reengage dormant users. However, this framework distinguishes itself from most of them by putting a lot of priority on how the personalization and predictive analytics can be applied in email campaign strategy management in a user-by-user basis rather than in a generic, generic manner. Following this methodology, marketers will be able to hang onto more users, enhance conversion rate and stimulate additional prolonged interaction in the platforms.

2. Literature Review

Cold start emails campaigns to re-engage dormant users in large scale platforms is an area of research in evolution. I found that there have been many examples of studies focusing on the difficulties and strategies to re-engage inactive users in the area of digital marketing and email outreach. The existing literature related to the use of data driven techniques such as time series analysis, predictive modeling, user segmentation, personalization and A/B testing for optimization of the email campaigns of large-scale platforms, in order to minimize cold start, is reviewed in this section.

2.1 The Challenge of Cold Start Email Campaigns

Email campaigns aimed at cold start target inactive or dormant users who have been away for extended periods of time, which is a fundamental feature of cold start. Dormant users do not have an up-to-date engagement history and therefore would be a difficult audience to personalize content and predict user behavior on through traditional email marketing strategies. Based on the campaign's cold start, researchers have discovered that the engagement signals are not enough to craft such campaigns, instead, they should find the way to leverage historical data as well as use advanced analytics [2]. Reacting to reengage these users is difficult, much less figuring out which tactics (timing, content, offers) will prompt them to take action. The fact of the matter is that the complexity of the problem suggests that we need a data driven technique, that can adaptively change strategies depending on user behaviors and preferences.

2.2 Data-Driven Approaches to Cold Start Campaigns

Successful cold start email campaigns are driven through data. With the volume of millions of users coming on huge platforms, it becomes essential to do a comprehensive analysis of how users behave, interact, and overall, what have they been engaging so far to make that targeted campaign. In marketing field, time series analysis has been extensively adopted with an aim to track and make forecasts on user engagement over time. Platforms can use this pattern of user activity to determine what the trends of when users are most likely to re-engage, and therefore how we might best time email campaigns to reach users to maximum potential. For instance, the emails of dormant users can be predicted using autoregressive integrated moving average (ARIMA methods) to forecast when these dormant users are most likely to open emails and engage with them [4]. Moreover, this study explains how trends on past behaviors (for example, frequency of email opens, time on the platform) are helping predict the future engagement and thus aid in more accurate, timelier outreach.

2.3 User Segmentation and Behavioral Targeting

For cold start campaigns, segmentation is an important component to the re engagement strategy. Different users have different reasons for inactivity and dormant users are not a homogenous group. Thereby, the past behaviors of the users, demographics and engagement history of the users helps to segment them and hence for more specific outreach. A number of studies have looked into segmentation methods that identify the dormant set of the users with the highest potential for re-engagement. As an example, recommend leveraging RFM (recency, frequency, monetary) [5] models to divide the users into different groups according to how long ago they rely on the platform, how frequently they do so, and how much they value their interactions. Particularly, segmentation allows lowering the level of efforts on users who are likely to react to targeted outreach increasing cold start emails efficiency.

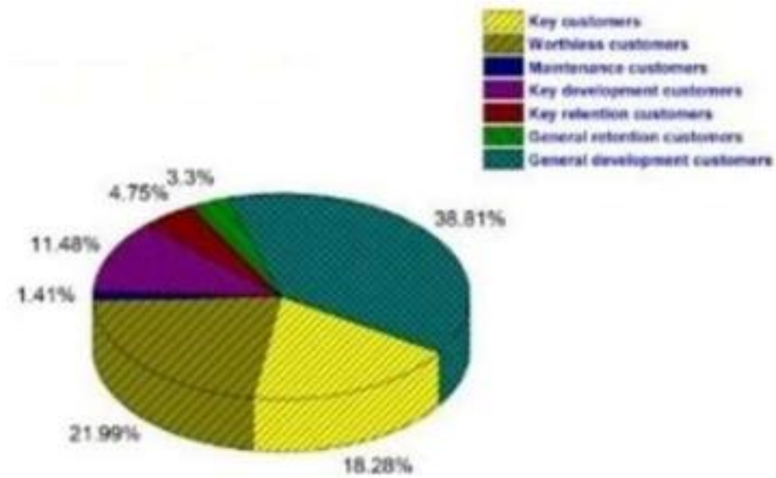

Figure 2. RFC Clustering [5]

Expand the research by applying machine learning techniques, including clustering algorithms, to weigh how well different users can be grouped by their characteristics and then predicting which group is most likely to re-engage. This can show marketers, based on the segmentation, to whom you can send the personalized email content, e.g. people who previously bought items may be offered discounts and people who used the content based you can re-engage with personalized recommendations. Given that timing and relevance are important to the success of cold start campaigns, personalization is critical and segmentation helps to ensure personalization.

2.4 Predictive Analytics for Re-Engagement

Thus, predictive analytics has been instrumental in the prediction of the dormant users having the highest probability of re-engaging with email campaign. User reengagement prediction based on historical data has been done using some of machine learning techniques such as regression models and neural networks. Predictive models measure user behavior from past, like email interaction, browsing behaviors, and purchasing history to speculate the probability value of re engagement. The predictive models such as logistic regression and support vector machines (SVMs) can be used to predict user responses to cold start campaigns with large datasets [6]. According to their findings, predictive models can be used to uncover important factors like open rates of emails and previous engagement that have a major effect on the effectiveness of reengagement efforts. Also, deep learning techniques are proven to be a good fit to recognize patterns within user behavior through time, using networks with Long Short-Term Memory (LSTM). Using these models, dormant user histories can be analyzed over a long time period and one can predict when dormant users are going to return the most. For example, series data such as the past interactions inform the future outcomes making LSTM networks an effective way to model such data. Adding the time element to predictive models allows platforms to time their outreach as closely as possible and most effectively modify their email content for the highest chance of re-engagement.

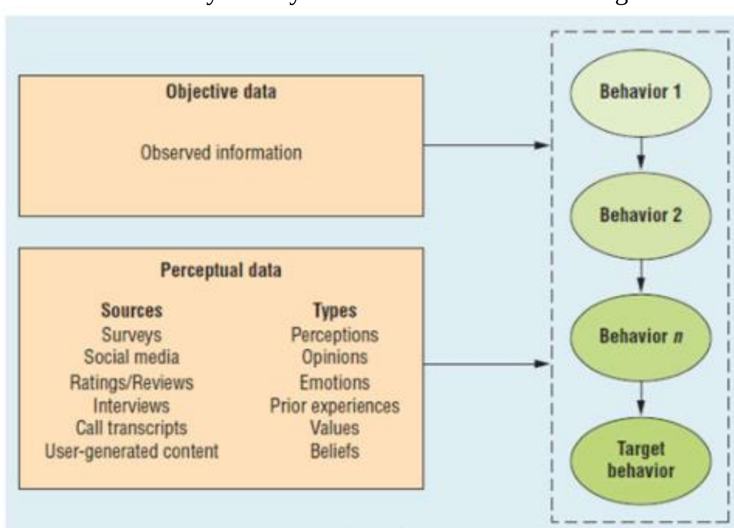

Figure 3. ARIMA [6]

3. Methodology

The methodology describes research strategies to develop a data-driven approach for running email campaigns at large scale platforms targeting dormant users through data collection methods and analytical approaches and evaluation metrics. This research applies a mix of data analytics and machine learning models as well as predictive analytics in creating and optimizing email campaigns to have a personalized strategy in reengagement. The research methodology consists of five essential components which begin with research design followed by data collection then data processing and model development before evaluation takes place.

3.1 Research Design

An effective framework to re-activate inactive users using cold start email campaigns is developed using quantitative research design with machine learning and data analytics. Predictive modelling forms the main goal of this project alongside comprehending user behavior patterns for developing forecasting systems that anticipate re-engagement behaviors. This is applied in a hybrid manner, where we incorporate actions by combining historical user data that is real time based and other engagement signals.

The implemented framework requires these sequential steps for execution:

- The system identifies dormant users when they stop using the platform for specific time lengths ranging from 30 to 90 days. The classification depends on evaluating email interaction records together with login events and purchase history and platform content engagement metrics.
- Dormant user segmentation creates distinct groups using combination methods of historical data together with demographic statistics and past engagement data.
- The email campaigns contain personalized text with specific offers targeted at different segments based on their historical engagement history.
- Machine learning algorithms are applied in Predictive Modeling to predict how likely a user would be re-engaged by analyzing different user and environmental parameters.

3.2 Data Collection

This research utilizes the user database of a major platform to obtain its data. The gathered data comprises user activity records from the past as well as interaction logs that reveal important patterns in user behavior. Key data points include:

- The demographic attributes of consumers that need segmentation include their geographic locations and their age along with additional relevant personal details.
- The complete data record of users' responses to received emails includes open rate statistics as well as click-through rate and conversion performance metrics.
- Users participate on the platform through their login behavior and downloading content in addition to utilizing platform features and spending time on the site.
- The platform records all buying activity and monetary interactions between users together with other financial deals that take place.
- Seasonal variables, daily times and external environmental conditions that could impact the way users engage with the platform are classified under environmental factors.

Internal tracking systems from the platform provide the collected information after all personal identifiers get removed to meet privacy and ethical requirements.

3.3 Data Processing and Feature Engineering

The raw data processing step makes data ready for analysis and model development purposes. Several critical functions are included in this processing stage.

- Data Cleaning includes identifying and removing duplicate records while handling empty spaces and unusual points which would otherwise affect evaluation results.
- The method selects essential characteristics from user mail records and system usage patterns and content-related interactions which affect reactivation behaviors. The exploratory data analysis (EDA) in combination with domain expertise generates features for use in the project.
- Time-Series Transformation: If the users have low activity, we try to analyze trends of long-term engagement and seasonality through the time series analysis methods such as resampling and smoothing.
- The normalization technique scales numerical features including login numbers and total spending amounts to achieve uniformity values throughout the dataset.

During model development and validation, the processed information divides into training and testing sections.

3.4 Model Development

Multiple machine learning approaches are used for developing models that forecast user reactivation. Both structured information consisting of demographics along with engagement background and sequential time-based data can be processed using these models. The following models are employed:

- **Clustering Algorithms for Segmentation:** Users can be segmented according to their past engagement methods through K-means and DBSCAN clustering operations. The tailored marketing endeavors utilize these segments because they sort users into homogeneous customer clusters based on their common traits. Hierarchical clustering generates results that study how users behave during dormancy stages and interact at different levels.
- **Predictive Models for Re-Engagement:** Logistic Regression together with Random Forest models perform to calculate re-engagement probabilities. The predictive models absorb characteristics from both users and their specific campaigns through features such as demographics and past interaction as well as email timing and content type. To predict the re-engagement probability in case of time series data, we use Long Short Time Memory (LSTM) network with sequential user behavior (e.g. frequency of past email opens, recency of logins).
- **Survival Analysis:** Using Cox Proportional Hazard Model enables the estimation of re-engagement duration for dynamic control of email outreach timing.

Both training and validation of the models happens using the training dataset and testing dataset respectively. The model performance reaches its peak after applying grid search and cross-validation during the hyperparameter tuning process.

3.5 Campaign Design and Personalization

The predictive models guide the development process of cold start email campaigns after training completes. The campaign format for each user segment focuses on increasing re-engagement opportunities throughout its design process. The campaign design incorporates these main features:

- Customers receive customized subject lines which leverage their recorded interaction records (such as viewed products and interacted content) to boost open rate performance.
- The email system tailors its content specifics to user preferences alongside their previous email interactions. Users who have made previous purchases will receive discount promotions yet those who interact with content will obtain customized content suggestions.
- Model predictions enable selecting the most suitable moment to transmit the email while determining the correct number of subsequent email deliveries.

The automated email marketing software together with predictive models implements the email creation schedule process.

3.6 Evaluation and Metrics

The evaluation of cold start email campaign effectiveness uses multiple key performance indicators (KPIs) for assessment.

- Open Rate represents the total quantity of users who open the sent email A high opening percentage shows the subject line together with timing worked well for the recipients.
- Users who press links inside email messages represent the Click-Through Rate (CTR). When email content proves engaging and relevant to subscribers the CTR values increase.
- The desired platform action completion rate by users defines conversion rate. Acquiring data about the re-engagement success of campaigns depends on this performance indicator as a fundamental measure.
- The percentage of idle users who come back to the platform becomes known as Re-Engagement Rate. The cold start campaign success depends primarily on achieving this metric as its main evaluation factor.
- The survival analysis technique measures how long users need until they return to the system after being sent emails.

The results of A/B Testing determine optimal email campaign variations through element comparison analysis between subject lines and promotional content types.

3.7 Ethical Considerations

This study adheres to ethical guidelines for handling user data. Protection of user privacy is ensured by removing identifying information then researchers obtain necessary consent from users. The research undergoes checks to obey both GDPR data protection requirements and ethical criteria in marketing research.

4. Results And Discussion

Generally, the predictive power of the machine learning models, for example, Long Short-Term Memory (LSTM) network, have been shown to be effective for improving user engagement strategy in various platforms, especially those from email campaigns' optimization. We use insights from recent research to distil three key observations and comment on how they pertain to cold start email re-engagement frameworks.

4.1 Predictive Superiority of LSTM Models in User Engagement

In Sarkar and De Bruyn (2021) [1], they conducted a thorough comparison between traditional models which needed the labor involved in manual feature engineering, versus LSTM based deep learning models for task of direct marketing. On the basis of 271 engineered models for customer behavior prediction (i.e., churn, brand choice), the LSTM network performed better than 269 of them. Thus, it clearly reveals that LSTM models have the ability to learn complicated temporal relationships among users' behavior without the need of extensive human preprocessing. This characteristic makes LSTMs, particularly useful for so-called cold start campaigns where one does not have much historical engagement data (or worse, may even be noisy), as they can learn representations directly from raw event streams.

4.2 Modeling Email Activity with Temporal Learning Models

The use of LSTM and RNN models in studying email traffic modelling (2020) [1] only supports the applicability of these models in our case. When these two models are trained on large email traffic datasets from multiple university servers, the authors showed that they achieved high accuracy. Temporal patterns in email activity could be successfully predicted by the models and benefited from an order of magnitude higher reliability than traditional probabilistic models. This is a validation for applying LSTM models in the cold start email re-engagement settings where awareness of the cadence and the clustering of the email interactions are essential to time and target dormant users.

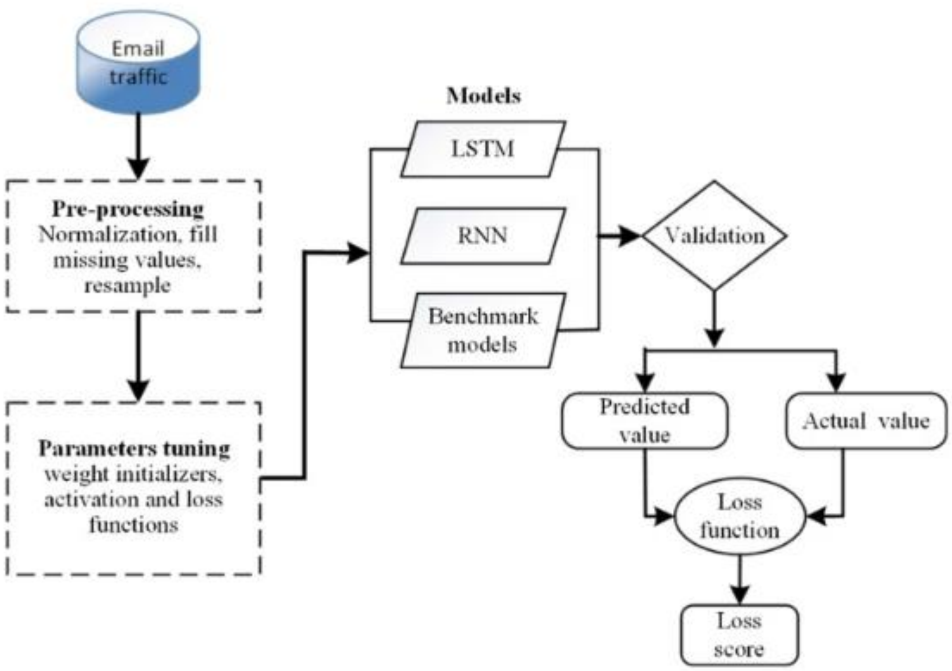

Figure 4. Model Architecture [2]

4.3 Importance of Context-Aware Temporal Predictions

Peters et al. (2023) [1] extend the LSTM modeling to the prediction of social media engagement using Snapchat user sessions comprising of over 100 million users. Beyond showing that past behavioral data is sufficient to predict user engagement (active, or posting, as well as passive, or viewing), we show that making better use of this past data is critical for predicting user engagement well. What they also found is that having the signal sent from a trusted environment (known as contextual signal) enhanced the predictive performance of their LSTM model significantly. Although the application is in a social media context the findings are amazingly applicable to large-scale email platforms when contextual signals (such as the user's last active time, device type, or campaign exposure) can be used to enhance targeting logic.

4.4 Implications for Cold Start Re-Engagement Strategy

Taken together, these studies have three major implications about how to try to reenable dormant users via an email.

- *LSTM Based Systems Do Not Require Much Data Processing:* This is because LSTM based systems can work on raw engagement signals and adapt to these to a great extent.
- The ability of these models to study temporal interaction patterns from data elements such as email reading timestamps enables them to identify the best times for reactivation.
- *Contextual Engagement:* Specially enhancing the real time contextual data as input to LSTM will make a huge difference in the outcomes of email engagement and provide email content and timing personalization.

In short, the results of existing empirical work are that compositional time series analysis and deep learning are feasible for developing adaptive intelligent cold start email strategies for reactivating dormant users.

5. Challenges and Considerations

Data-based approaches to user reactivation through cold start emails show promise but organizations need to handle numerous practical and technological barriers to achieve large-scale success. The following segment details important organizational matters which need attention to develop effective ethical cold start mail marketing initiatives.

5.1 Data Sparsity and Cold Start Limitations

The major obstacle during cold start campaigns emerges from having no behavioral information regarding inactive users. The shortage of interaction data makes the process of generating proper customer profiles and predicting user reactivation times very problematic. The scarce amount of data affects the development of strong machine learning models which produces possible errors in customer segmentation and targeting systems. The resolution of this issue requires organizations to utilize peer group behavior indicators together with past partial interaction records and contextual data signals.

5.2 User Privacy and Ethical Constraints

Behavioral analytics in email marketing requires organizations to fulfil GDPR and CCPA user data protection rules. The practice of processing behavioral profiles through inference poses significant problems for user consent regulations and corresponding ethical as well as legal concerns. Users must receive comprehensive data transparency alongside options to remain anonymous while accessing data protection in order to create systems based on respect for their privacy according to established standards. The perception of a brand becomes adversely affected by targeted marketing approaches that extend beyond normal boundaries.

5.3 Model Interpretability and Bias

Users with complex LSTMs often generate reliable outcome predictions yet these systems produce results without explanation capabilities that marketing teams must understand about user groups and content strategy development. Such models will automatically acquire biases from unbalanced data collections including specific demographic groups or activity types thus resulting in substandard or discriminatory re-engagement methods. These problems call for both explainable models and ML systems that respect fairness principles.

5.4 Email Deliverability and Fatigue

Set Email re-activation of dormant users depends on the accurate predictions generated by predictive models as well as reliable deliverability frameworks. Email effectiveness suffers greatly when recipients exhibit high bounce rates and implement spam philters in combination with email fatigue. Dormant users usually do not engage with bulk messages when they receive them repeatedly and the attempts seem unconnected to their past interactions. The campaign design required embedded control systems for timing purposes alongside relevant content preparation and frequency calculations.

5.5 Real-time Scaling and System Integration

A large-scale platform must merge predictive models with existing CRM and email automation and analytics pipelines for the deployment of its re-engagement framework. Technical challenges stemming from real-time performance and system update synchronization affect efficiency together with user experience while minimization of prediction-to-campaign launch latency represents an additional challenge. The infrastructure design process needs to place highest importance on features for scalability and fault-tolerance as well as low-latency decision capabilities.

6. Future Directions

Multiple research and development approaches must be investigated to increase both the effectiveness and adaptability of data-driven frameworks that target dormant users. These directions seek to fill current operational deficiencies by boosting the predictive capabilities and customization level and ethical aspects of cold start email initiatives within big platforms.

6.1 Integrating Multi-Channel Engagement Signals

Future research should expand this investigation by combining email engagement data with activity notice from various platforms such as social media platforms and website interactions and mobile application notifications and support channels. Platner build unified profiles with information from multiple sources so they can better determine user interests and reaction levels. The concept of holistic viewing enables engagement strategies which become specific to known communication channels and personal interaction styles for each user.

6.2 The implementation of reinforcement learning operates in real time for personalization purposes.

With a dynamic, self-improving models, reinforcement learning (RL) is a promising opportunity to evolve from static to re-engagement strategies. Future computational frameworks should integrate RL systems to use instant user feedback for ongoing improvements of email message composition and delivery times and recipient subject line options. Such measures would let marketing efforts adapt to evolving user behavior while sending messages that have better reactivation chances and a lower chance of spam behavior toward inactive customers.

6.3 The use of generative models

Allows behavioral segmentation to take place. Limited behavioral data serves as the main cause of cold start problems. Here, the idea would be to study how to simulate user behavior that is missing with generative models such as Variational Autoencoders (VAEs) or Transformer architectures to do more accurate segmentation and targeting. The models detect hidden behavioral patterns within groups of similar users to forecast how customers will engage with the platform and create personalized content for dormant accounts since they either disappeared or failed to participate fully.

6.4 Ethical AI and Explainability in Re-Engagement Decisions

Machine learning serves as an essential core element in customer reactivation strategies so transparency in decision-making processes becomes essential to implement. Future AI frameworks must integrate explainable AI (XAI) systems which enable platform managers and marketing experts to understand their user selection processes for re-engagement together with the methods used to select content. By following this approach trust between organizations and their users increases while data protection regulations (such as GDPR) get respected by providing user-driven autonomy and preference awareness.

6.5 Longitudinal Impact Studies of Re-Engagement

The majority of studies about marketing campaigns examine only immediate results including open rate statistics and click-through rates and user login successes. Future investigations need to analyze the extended effects which develop from re-engagement promotional initiatives. Research should reveal how users who return to use the system behave across long periods including their progression from activity to customer retention and ultimate high-value customer status. Investing in long-term tracking will lead to better sustainable growth plans instead of promoting single-point connectivity enhancement [10].

6.6 Federated and Privacy-Preserving Learning

Future implementations are being anticipated to be with federated learning and differential privacy techniques as data privacy and the growing regulations of globally are becoming a concern. Decentralized data can be utilized for training machine learning models through these approaches which protect the confidentiality of sensitive information at all times. Using these methods to activate new campaigns allows businesses to follow privacy rules and analyze user patterns in aggregate form which protects both security standards and reputation.

7. Conclusion

A data-centered framework serves as the basis for this study to re-activate dormant users by using initial email campaigns across significant platform systems. Aligning with this idea, the proposed framework also takes advantage of compositional time series analysis and machine learning models alongside stratified segmentation to come up with a systematic way of understating dormant user behavior and maximizing email marketing results. It's possible of business that involves predictive analytics together with the continual personalization techniques could both increase user engagement and also help maintain user departure while achieving the long-term user bond. The marketing campaigns enjoy real-time adjustments through multi-channel integrated signals to deliver purposeful content to users by platforms. The method has demonstrated successful outcomes while facing three major challenges that relate to solving sparse data complexity and meeting privacy requirements and creating personalized solutions applicable to all consumer segments. Future developments of this method will integrate federated learning with reinforcement learning as well as longitudinal study implementation. Smart ethical strategies to protect user privacy need urgent development since big platforms have to maintain user activation against market changes.

References

[1] V. S. D. &. W. M. Fast, "Data-driven competitive advantages in digital markets: An overview of data value and facilitati factors.," 2021.

[2] M. W. G. K. K. M. C. F. G. S. S. &. U. S. Golec, " Cold start latency in serverless computing: A systematic review, taxonom and future directions.," *ACM Computing Surveys,* vol. 57, no. 3, pp. 1-36, 2024.

[3] B. A.-F. A. G. A. B. D. A. S. Q. J. &. F. A. C. Qolomany, " Leveraging machine learning and big data for smart buildings: comprehensive survey.," *IEEE access, 7, 90316-90356.*, 2019.

[4] S. Bostanshirin, "Online marketing: challenges and opportunities," *In ProceediInternational Conference on Social Sciences a Humanities, Istanbul, September,* pp. 8-10, 2014.

[5] A. J. V. K. S. D. Y. &. P. P. M. Date, "A Systematic Review of Evaluating the Efficiency of Different Models for the Task Customer Segmentation.," 2024.

[6] K. &. G. B. P. Bhatia, "Predictive Analytics for Customer Behavior and Sales Forecasting in Retail.," *International Journal Web of Multidisciplinary Studies*, vol. 2, no. 1, pp. 24-30., 2025.

[7] M. &. D. B. A. Sarkar, "LSTM response models for direct marketing analytics: Replacing feature engineering with de learning.," *Journal of Interactive Marketing,* vol. 53, no. 1, pp. 80-95., 2021.

[8] K. B. S. N. A. M. T. D. M. K. P. &. W. K. W. Om, "Modelling email traffic workloads with RNN and LSTM models.," *Huma centric Computing and Information Sciences,* vol. 10, pp. 1-16, 2020.

[9] H. L. Y. B. F. B. R. A. M. S. C. &. B. M. W. Peters, "Context-aware prediction of user engagement on online social platforms *arXiv preprint arXiv:2310.14533.*, 2023.

[10] M. B. B. E. L. &. H. J. G. Altier, "On Re-engagement and Risk Factors.," *Terrorism and Political Violence,* vol. 33, no. 4, p 868-874., 2021.