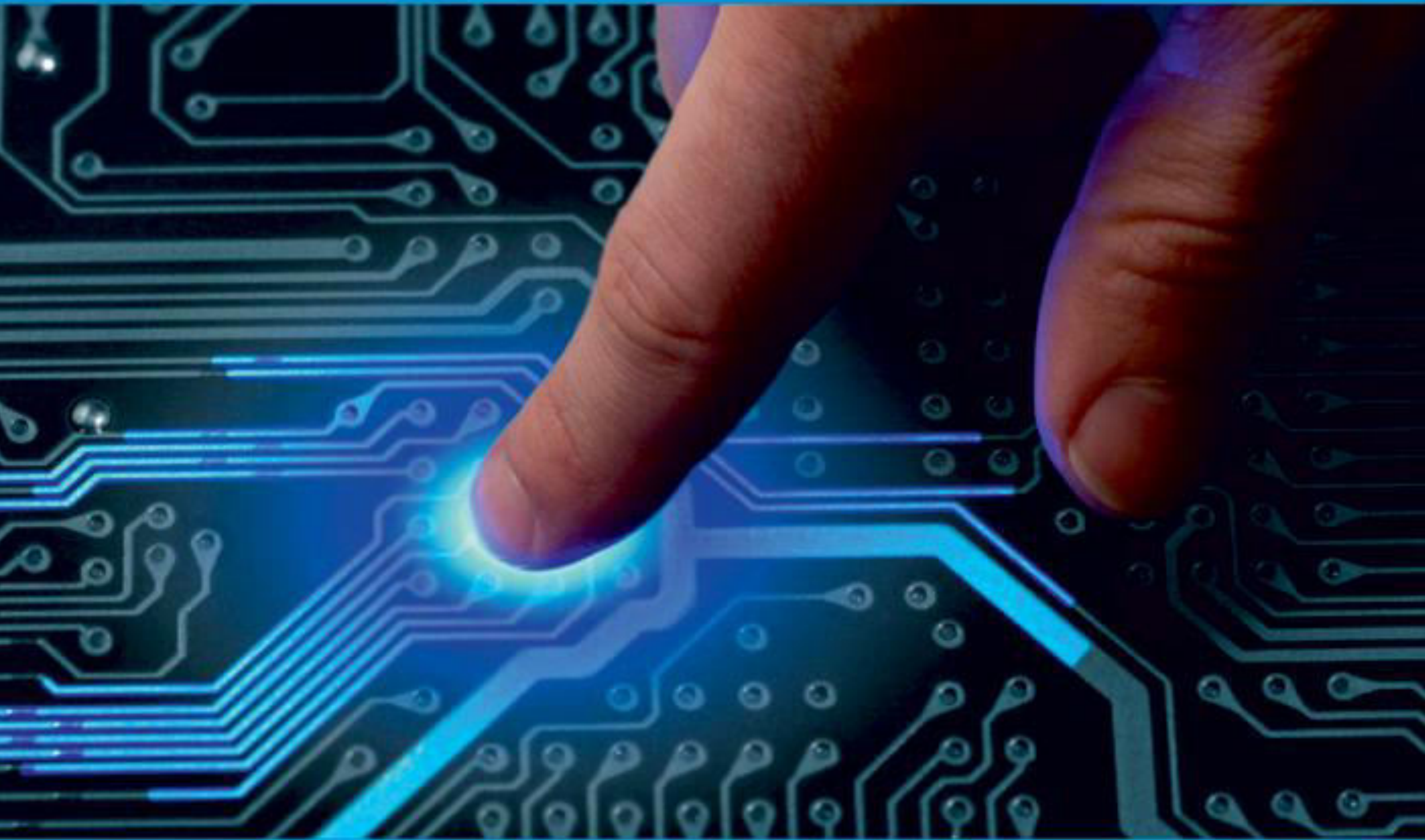
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 2, February 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

Artificial Intelligence Approach to Predicting Customer Churn in Social-Media Communities

Anusha Musunuri

Reddit Inc, California - USA

ABSTRACT: The rapid growth of social media platforms such as Instagram, Facebook, and Twitter has transformed how businesses interact with their customers. Rather than relying solely on traditional support or marketing channels, many companies now engage with users through online social-media communities to build brand loyalty, provide real-time support, and foster customer engagement. These communities, while beneficial, can also reflect early signs of dissatisfaction when users express frustration or reduce their activity. As a result, the ability to predict customer churn in these environments has become increasingly valuable for organizations seeking to improve retention and customer lifetime value.

Conventional churn prediction methods typically involve the manual analysis of structured customer data, such as usage frequency, subscription duration, or transaction history. However, these approaches often fall short when applied to the dynamic and unstructured nature of social-media interactions. The informal language, rapid content flow, and volume of user-generated data require more advanced techniques for meaningful analysis.

Artificial intelligence offers a more effective solution by leveraging natural language processing, sentiment analysis, and machine learning to extract insights from posts, comments, and engagement behaviors. AI systems can detect patterns that precede customer disengagement, such as declining interaction levels, negative sentiment, or abrupt changes in behavior. These insights allow businesses to identify at-risk users with greater accuracy and respond with timely interventions, including personalized outreach, targeted content, or incentive programs.

By integrating AI-driven analysis into their customer retention strategies, businesses can enhance their understanding of user behavior in social-media communities and take proactive measures to reduce churn. This approach not only preserves customer relationships but also contributes to long-term business sustainability in an increasingly digital and competitive marketplace.

KEYWORDS: Artificial Intelligence, Customer Retention, Subscription Models, Predictive Analytics, Churn Forecasting, Online Communities, Social-media

I. INTRODUCTION

Customer churn refers to the likelihood that a user will discontinue their participation in a product, service, or platform. In social-media communities, churn can manifest as a drop in engagement, reduced interaction, or complete withdrawal from the group. Identifying and predicting this behavior is critical for community managers and moderators, as it allows them to take informed steps to retain users before disengagement becomes permanent.

The data required for churn prediction in social-media environments is typically gathered from user activity logs, engagement frequency, time spent within the community, membership duration, and demographic attributes. After this data is collected, it is pre-processed through cleaning, normalization, and feature extraction to ensure quality and consistency. Once prepared, the dataset can be used to train machine learning models capable of detecting early signs of churn.

The choice of machine learning algorithm depends on the structure and purpose of the community. Since user behavior can differ significantly across various communities, no single model fits all scenarios. Commonly used algorithms include logistic regression, decision trees, random forests, and neural networks. These models are trained on historical data where the churn status is known, enabling the model to learn patterns that often precede user departure.

After training, the model is evaluated on a separate validation set to measure its performance using metrics such as accuracy, precision, recall, and F1 score. If performance is unsatisfactory, further tuning or model adjustments may be needed. Once validated, the model can be applied to current user data to generate real-time churn predictions. This

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

allows community managers to intervene early with targeted engagement strategies such as personalized content, rewards, or direct communication.

Predicting churn in social-media communities presents unique challenges. One issue is the ambiguity in defining churn. A user may become inactive in one subgroup while remaining active in others, which complicates churn labeling. Additionally, user behavior in these communities is highly volatile and can shift rapidly in response to external factors such as trends, platform changes, or real-world events. This variability reduces the reliability of static behavioral models.

Another critical consideration is ethical data usage. Since social-media platforms are often public or semi-public, collecting and analyzing user-generated content must be done with caution, adhering to privacy standards and user consent where applicable.

This research contributes in three key areas:

- **Correlation Analysis:**
  The study identifies key behavioral and emotional indicators associated with churn, including changes in engagement frequency, sentiment trends, and time since last interaction.

- **AI-Based Churn Prediction Model:**
  A machine learning model is proposed to predict churn using a combination of user activity metrics and sentiment analysis. The model provides early detection of at-risk users with high precision.

- **Practical Implications for Community Management:**
  The findings offer actionable insights for community managers, enabling them to proactively address potential churn through targeted outreach and community design improvements.

By integrating artificial intelligence into community engagement strategies, organizations can maintain stronger, more loyal user bases while responding adaptively to behavioral changes within their social ecosystems.

II. RELATED WORK

Kwon, Y. D.,et,al.[11] have discussed this problem, and we address the Geo Lifecycle of user engagement and churn prediction in location-based social networking apps. It meant monitoring user behavior and interaction on geographical features to infer long-term net retention. Spann, B.,et,al.[12] has discussed a machine learning problem, Predicting Toxicity, in online socia-media Discussion Threads 101. This filtering technique typically uses data analysis methods like algorithms to recognize and identify harmful or undesirable content within online messages. Wang, C.,et,al.[13] have discussed that customer churn prediction is extremely important in the internet funds industry so businesses can prevent holding on to customers and avoid losses. This study proposes a novel feature-embedding convolutional neural network-based methodology to predict customer churn with high precision. Extensive experimental results show the effectiveness of our proposed model in contrast to existing methods. Nayak, S.,et,al.[14] have discussed Machine Learning Algorithm will analyze our data sets and find patterns in online socia-media posts about mental health. It has the potential to illuminate discussions around specific mental health issues and discover any resources or interventions that are available for someone who is struggling with their mental state. Chang, Y. C.,et,al.[15] have discussed the study, which examines the potential to enhance next-generation decision support systems with aspects of personality assessment by reviewing new data from a large social media platform, online socia-media . This study aims to provide important findings that can help for more precise and successful recruitment using user behavior analysis and language. Nadiri, A.,et,al.[16] have discussed we analyze how long different types of users remain active on a social media site. Using user activity data for the analysis, this study seeks to identify what features drive how long a user stays present on their platform and see if it changes over time. Adelani, D. I.,et,al.[17] have discussed that feedback from the community is a key driver of what gets talked about on social media. Using historical trends and user activity, we can use predictive modeling to forecast how this feedback would influence topic selection. This is a valuable insight for marketers, helping them target and tailor their social media strategy. Thushari, P. D.,et,al.[18] have discussed the millions of social media interactions on online socia-media , algorithms from machine learning could process troves of data to find patterns and signal marks for mental health. Along with explainable AI methods that can give insights and explanations about how behavior, language, or sentiments affect mental health. Shah, F.,et,al.[19] have discussed using machine learning and natural language processing; artificial intelligence as a

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |
|| Volume 11, Issue 2, February 2023 ||
| DOI: 10.15680/IJIRCCE.2023.1102056 |

service is trained to identify immoral content, including hate speech, violence, and pornographic images on various web applications because it can process huge volumes of data at speed and learn over time to detect and remove this toxic content so that we have a better place on the web. Kilroy, D.,et,al.[20] have discussed using data from various product categories for analysis and making a prediction regarding the demand. By using algorithms to recognize patterns and trends in customer behavior, companies can predict future products or service demands so that they can adjust their service offerings on time. It makes enterprises more competitive and enables them to provide better services.

III. PROPOSED MODEL

The model devised for modeling the problem falls under Artificial Intelligence, where machine learning is a part of it. Its sub-field allows systems to learn through algorithms, models, and other mathematical techniques derived from previous data without being programmed explicitly. This process starts with data collection, as you will invariably need a vast and varied database for training your AI algorithm.

$$x_\theta(h) = \frac{1}{1+g^{-\theta V_h}} \tag{1}$$

$$C = \frac{2|G|}{|T|(|T|-1)} \tag{2}$$

This information can be in the form of sensor data, text, images, or videos. It pre-processes the data to clean and arrange it and extracts related features. They retrain the model with varying parameters or algorithms if the performance is poor. After training and testing your model, it can now be used in real-life applications. Given some new data, it can decide or predict, and the more you provide with emphasis on the quality of text, the better its performance will be over time.

$$C = \frac{|G|}{|T|(|T|-1)} \tag{3}$$

This AI approach is a flexible and learnable model that can improve over time. It can revolutionize numerous industries, including health care, finances, and transportation, by streamlining processes, recognizing patterns in data, and making well-thought-out decisions.

3. 1. Construction

Kiaty Driadefer defines artificial intelligence as the ability of computers or machines to perform human-like learning and cognitive functions. The process of designing, enacting, and performing algorithms for machines to collect data and analyze it based on pre-established rules or models in which decisions can be made from the analyses solving a particular problem. This information is the data that machines tap to learn and make decisions. Building any AI system requires huge amounts of data. This data can be of two types: either pre-labeled or the data being used for prediction-making, which has no labels and the algorithm, will learn from this data. The idea is to have an algorithm to digest the data, find trends, and make decisions based on those trends.

Classified Output: Pre-process the data, classifying it as churners and non-churners of the dataset. Churners are the customers who have left their usage of a product or service, while non-churners are those still classified.

Churners: Customers who have not engaged or stopped using your products/services. The customer's behavior can identify churners, whether a reduction in usage, missed payments, or a low engagement rate. Fig 1 Shows the Construction Model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

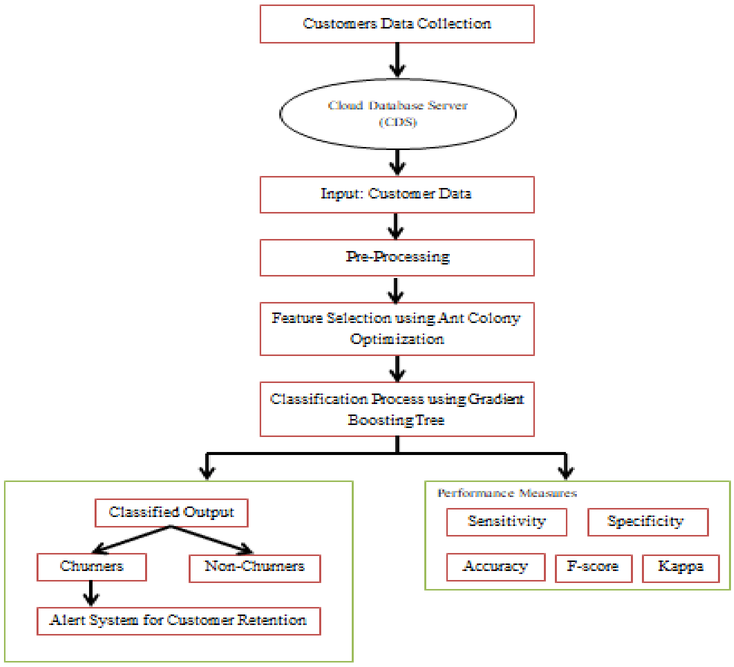


Fig 1: Construction Model

Performance Measures: We need performance measures to assess the predictive capabilities of churn prediction models. These measures are accuracy, specificity, sensitivity, and F1 score.

Sensitivity: This is also called recall, representing the proportion of churners our model identifies. Put differently, it quantifies a model's capability to label churners accurately.

The third component is computing power. The advent of technology has made it possible to design machines capable of rapidly dealing with huge quantities and types of data in less time.

$$\log it(f) = In\left(\frac{f}{1-f}\right) \tag{4}$$

This computing power is critical to allow AI systems to run complex algorithms and make instantaneous decisions. An AI system incorporates several processes, ranging from data collection and preprocessing to algorithm selection/analysis, followed by a model training/testing process. A group of experts from different backgrounds must collaborate to create a great AI system.

3. 2. Operating principle

Artificial Intelligence is a field of computer science that seeks to construct machines that can follow methods for human-like tasks. AI, the way AI is going about it, involves creating algorithms that can replicate human cognitive capabilities like learning, reasoning, problem-solving, and decision-making.

Initializing the parameters: Parameters used for Food source optimization algorithms are the number of food sources, maximum iterations number and employed honey bees, onlooker size category specified search space.

Leader Position: Leader Position is the best solution fitness value found so far in search space in the behavior of food source optimization algorithms. The first leader will be set at a random solution in the search space. Fig 2 Shows the Operating Principle Model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

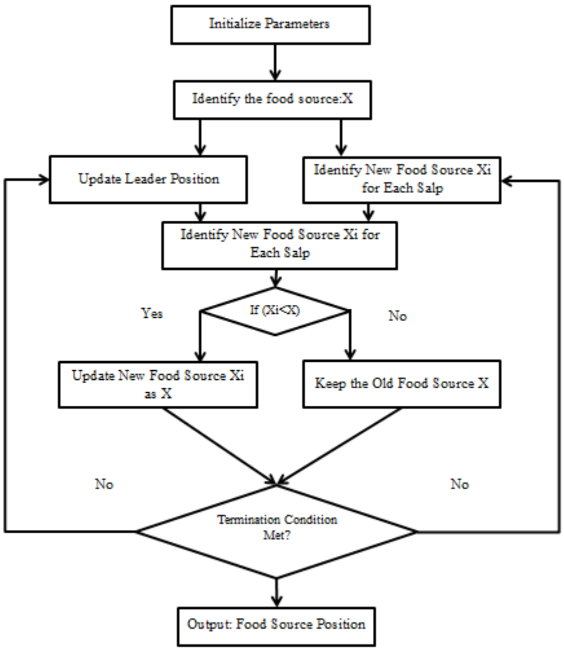


Fig 2: Operating Principle Model

Position of followers: Followers in food source optimization algorithms are known as employed and onlooker bees. Their positions are shaped relative to the leader's position and the food sources they patrol. The aim of the followers is to get solutions better than their previous location and also from the whole solution.

Food source: Food source in food source optimization algorithms denotes the candidate solutions of an optimization problem. They have been assigned to worker and observer bees based on a fitness function.

Machine learning is one of the main methods in AI, and it consists of training computer programs with lots of data to understand patterns that would aid them in making decisions. AI possesses the ability to process natural language by which machines can understand humans and respond to them. It requires analyzing and interpreting spoken or written language using statistical machine-learning techniques.

$$E=\sum_{k=1}^{Y} \hat{y}_{ny}\left(1-\hat{y}_{ny}\right) \tag{5}$$

AI operates on the principle of developing a technology that can perform tasks required by human intelligence, such as learning, decision-making, and language understanding

IV. RESULTS AND DISCUSSION

Prediction Accuracy: This parameter in this project gauges how well the AI approach can predict when a customer will leave various online socia-media communities. Fig 3 shows the Computation of prediction accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

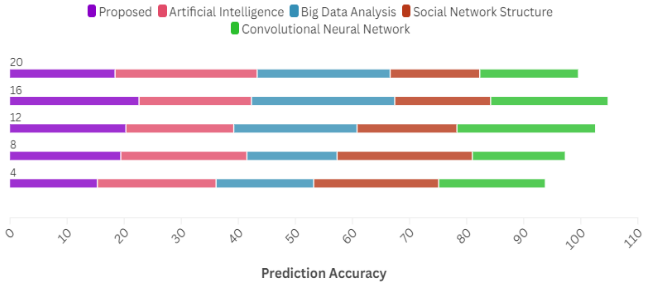


Fig 3: Computation of Prediction Accuracy

The higher the prediction accuracy, the better the AI Model identifies possible churners and the fewer incorrect predictions.

Computational Efficiency: This parameter quantifies how fast and effectively the AI model can process or analyze enormous datasets from online socia-media communities. Fig 4 shows the Computation of computational efficiency.

Fig 4: Computation of Computational Efficiency

We need a model that predicts in real-time and is fast enough to help prevent churn.

Data Quality: Training data quality is a performance dimension. Thus, the more precise this data is with relevant and accurate features, the better our churn prediction will be. Fig 5 shows the Computation of data quality.

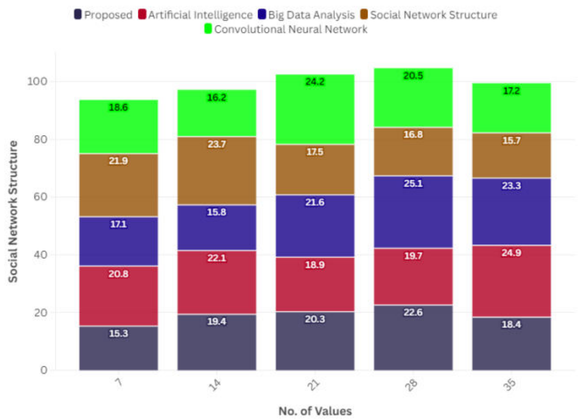


Fig 5: Computation of Data Quality

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

Low-quality data can contribute to biased predictions. Scalability: Scaling predictions across a growing data set and online socia-media communities is an important technical capability for any AI model. Fig 6 shows the Computation of miss rate.

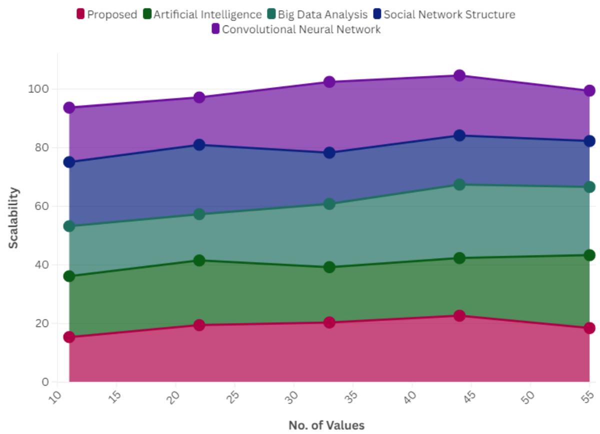


Fig 6: Computation of miss rate

The most scalable model is capable of supporting growth for YEARS without having to retrain, and it can handle very high-accuracy churn predictions while improving speed.

V. CONCLUSION

Customer Churn Prediction with AI in online socia-media 's Potential for Identifying Churners AI can then analyze huge amounts of data using machine learning algorithms and natural language processing techniques, identify behaviors that may lead to customer churn patterns in social media conversations, website navigation etc. It provides advantages like enhanced prediction performance, quicker computation and processing speed, and newly detected complex patterns containing unstructured data. Nonetheless, AI is not a panacea; it requires human involvement and domain expertise to be implemented successfully. Improvement of the technique is based on how fast AI–based technology is progressing.

REFERENCES

1. Wang, X., Zhao, K., Zhou, X., & Street, N. (2020). Predicting user posting activities in online health communities with deep learning. ACM Transactions on Management Information Systems (TMIS), 11(3), 1-15.
2. Almerekhi, H., Kwak, H., & Jansen, B. J. (2022). Investigating toxicity changes of cross-community online socia-media ors from 2 billion posts and comments. PeerJ Computer Science, 8, e1059.
3. Bergstrom, K., & Poor, N. (2022). Signaling the intent to change online communities: A case from a online socia-media gaming community. Social Media+ Society, 8(2), 20563051221096817.
4. Tsugawa, S., & Niida, S. (2021). Analyzing the effects of social network Structure on the growth and survival of online communities in online socia-media . IEICE Transactions on Communications, 104(7), 760-769.
5. Liu, D. R., Liao, H. Y., Chen, K. Y., & Chiu, Y. L. (2019). Churn prediction and social neighbour influences for different types of user groups in virtual worlds. Expert Systems, 36(3), e12384.
6. Sadeque, F., & Bethard, S. (2019). Predicting engagement in online social networks: Challenges and opportunities. arXiv preprint arXiv:1907.05442.
7. Kwon, Y. D., Chatzopoulos, D., ul Haq, E., Wong, R. C. W., & Hui, P. (2019). GeoLifecycle: User engagement of geographical exploration and churn prediction in LBSNs. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3), 1-29.
8. Spann, B., & Agarwal, N. (2023, September). Predicting Toxicity in online socia-media Discussion Threads. In In Proceedings of the 16th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2023).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102056 |

9. Wang, C., Han, D., Fan, W., & Liu, Q. (2019). Customer churn prediction with feature embedded convolutional neural network: An empirical study in the internet funds industry. International Journal of Computational Intelligence and Applications, 18(01), 1950003.
10. Nayak, S., Mahapatra, D., Chatterjee, R., Parida, S., & Dash, S. R. (2022). A Machine Learning Approach to Analyze Mental Health from online socia-media Posts. In Biologically Inspired Techniques in Many Criteria Decision Making: Proceedings of BITMDM 2021 (pp. 357-366). Singapore: Springer Nature Singapore.
11. Chang, Y. C., Liew, D. J., & Ku, C. H. Utilizing online socia-media Data for Personality Prediction to Enhance Job Fit Assessment in Decision Support Systems. Available at SSRN 4762187.
12. Nadiri, A., & Takes, F. W. (2022, April). A large-scale temporal analysis of user lifespan durability on the online socia-media social media platform. In Companion Proceedings of the Web Conference 2022 (pp. 677-685).
13. Adelani, D. I., Kobayashi, R., Weber, I., & Grabowicz, P. A. (2020). Estimating community feedback effect on topic choice in social media with predictive modeling. EPJ Data Science, 9(1), 25.
14. Shah, F., Anwar, A., ul haq, I., AlSalman, H., Hussain, S., & Al-Hadhrami, S. (2022). Artificial intelligence as a service for immoral content detection and eradication. Scientific Programming, 2022(1), 6825228.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING