

An Adaptive Classification Framework for Handling the Cold Start Problem in Case of News Items



Azhar Yousf Mir¹, Majid Zaman^{2*}, Syed Mohammad Khurshid Quadri³, Sheikh Amir Fayaz¹

¹ Department of Computer Sciences, University of Kashmir, J&K, Srinagar 190006, India

² Directorate of IT&SS, University of Kashmir, J&K, Srinagar 190006, India

³ Department of Computer Science, Jamia Millia Islamia, Delhi 110025, India

Corresponding Author Email: zamanmajid@gmail.com

<https://doi.org/10.18280/ria.360609>

ABSTRACT

Received: 20 November 2022

Accepted: 20 December 2022

Keywords:

cold start problem, recommendation system, ID3 algorithm, browsing behavior, location based system, news recommendation

Many recommendation systems make product or service suggestions based on existing knowledge of the user or the item. We must deal with two types of cold start problems: item-based and user-based cold start problems. In a user-based cold start dilemma, it is difficult for the system to suggest news to a new user whose information is not saved in the system. In this article, we attempt to address user based cold start problem by assuming that in the case of a user, we only know one type of information about the user, and that information is the user's location. Using BBC news data, an ID3 classification approach was developed, which incorporates eight explanatory factors such as News ID, News text, Keywords, date, Location, Shares count, followers, and so on. The classification accuracy of one of the best fit models constructed using (80-20)% training and test ratios is around 78%. Our technique is an effective tool for the cold-start problem because it outperforms the advice by a significant margin depending on the location. According to the results, our approach is competitive in terms of both accuracy and precision.

1. INTRODUCTION

Recommender systems are software tools that describe user preferences in order to recommend goods to consume. Recommender systems are extremely important in news consumption. Where recommendations are created by processing massive amounts of data in order to offer new content that best fit user tastes. The technique of rating estimate used by recommender systems is typically categorized. We may distinguish between content-based and collaborative recommender systems in general. The user often bases content-based suggestions on item resemblance to previously liked things. In contrast, collaborative recommendation systems rely on ratings provided by people who share similar tastes and preferences [1]. However, each strategy has its own set of downsides. The main issues with content-based recommender systems include restricted content analysis, overspecialization, and the new user problem. Content-based (CB) strategies primarily examine item properties that were derived automatically by information retrieval methods. While powerful algorithms exist to analyze textual texts, feature extraction may be far more difficult to apply to multimedia data or things with diverse properties [2]. Furthermore, content-based recommendations overspecialize since only products with a high resemblance to those already evaluated are offered to the specific user.

Customers always begin their online shopping experience by browsing for the things they want. A parent might expect to seek for toys for their children, whereas a teenager could expect to look for new games, technology, mobile phones, course lessons, and so on [3, 4]. The recommendation system, by definition, is used as a therapy to help in the choosing of

various goods. It is beneficial in a variety of sectors, including e-commerce platforms such as Amazon, Flipkart, and others. Furthermore, it has shown its viability on travel-related websites such as TripAdvisor, MakeMyTrip, and others [5, 6]. A system like this is based on product attributes, user profiles built on websites, and product-related data. The most challenging part is integrating user information with a specific object.

Collaborative filtering (CF) recommender systems, as opposed to content-based recommendation approaches, create predictions based on things already evaluated by other users. Nonetheless, collaborative recommender systems display the new user issue, and in order to produce trustworthy suggestions, they must first learn user preferences. Aside from the new user cold start problem, collaborative systems also display the new item problem, which implies that a new item must be assessed by a sufficient number of users in order for the system to propose it properly [7]. Furthermore, the amount of rating information accessible determines the success of a collaborative recommender system. In most cases, the number of ratings acquired is insignificant in comparison to the number of ratings that must be projected. In other words, the user-item matrix is typically relatively sparse, resulting in poor suggestions.

Another sort of recommendation system is the Hybrid Recommendation System, which combines the benefits of CF and CB while minimizing their shortcomings. Basic recommender systems such as Content-Based, Collaborative Filtering, and Demographic recommender systems are combined to form hybrid recommender systems [8, 9]. The main challenges with basic recommender systems are cold-start (in CB, CF, and demographic) and stability/plasticity

concerns (where it is difficult to adjust a user's existing profile, i.e. user may want to tick things that are different from its preferred trend).

Readers who browse online news sites may benefit from customized news recommendations to prevent information overload. Customized news suggestion algorithms in use frequently match the substance of candidate news pieces and user interest as determined by previous activity today [10-13]. For example, authors proposed modelling news headers after several heads of self-attention. In order to capture the linkages between diverse activities, they employed multi-head self concentration to model user interest from previously clicked news items. The author proposed utilizing the CNN network to predict both long-term and short-term user interests based on news click behaviour and to learn news embedding's based on news headlines and categories [14]. These tailored news recommendation algorithms often fail to deliver credible suggestions to these users due to the scarce activity of cold-start users and the difficulties in modelling their interests [15, 16]. Furthermore, these algorithms commonly recommend items that users have already read which may reduce user interest and hinder users from acquiring new information.

Numerous recommendation systems base their suggestions for items or services on the user's or the item's prior knowledge. Cold start issues of two different types—item-based and user-based—must be addressed. It is challenging for the system to recommend news to a new user whose information is not saved in the system in a user-based cold start issue. The key contribution of this research is our attempt to handle the user-based cold start problem under the assumption that we only have access to one sort of information about a user, namely their location.

The paper is organised as follows: The literature on cold start issues and the numerous studies offered by the many researchers are briefly outlined in Section 2. Section 3 discusses the general approach of the recommendation system and current issues. Section 4 describes the experimental design, data collection, and preparation of the data used in this study. Section 5 specifies experimental performance. The entire process of the study is covered in Section 6, and Section 7 offers some recommendations for the future and concludes the study.

2. LITERATURE SURVEY

The cold-start issue the term "recommendation" can refer to either new things or new users, and we will focus on the latter. Some solutions to the problem necessitate an initial, albeit minor, set of preferences regarding the newcomer user [17]. The recommender system has a decade-long track record of increasing commercial values for businesses. Famous recommender system approaches such as collaborative filtering, matrix factorization, learning to rank, and deep-learning based recommendation [18, 19] have all been used extensively in businesses throughout the world. Many researchers have proposed their solutions and challenges they faced and some of these are scripted below:

A study [20] was proposed in which embedding-based news recommendation for millions of users was taken into account and an auto-encoder learned news embedding's from news bodies, while a GRU network modelled user interests from clicked stories was used. The dot product of their embedding's defines the match between the news and the user.

The authors of this study [21] handle the cold-start problem by promoting content that no one in the community has seen before to any new users who have no saved preferences. While there have been numerous research to remedy the cold start problem, they only addressed item-cold start or user-cold start, and the solutions still suffer from the privacy issue. As a result, the authors created a privacy-protected paradigm to handle the cold start problem (both user and item cold start). The authors proposed two types of recommendations (node recommendations and batch recommendations), and they compared the proposed method to three other alternative methods (Triadic Aspect Method, Nave Filterbots Method, and MediaScout Stereotype Method), and the dataset used was collected from online web news to generate recommendations using their method and the other three alternative methods. They computed the degree of novelty, coverage, and accuracy. In comparison to these three techniques, they discovered that our method obtained a better degree of novelty in batch recommendations while also achieving higher levels of coverage and precision in node suggestions.

In this paper [22], the authors propose ZeroMat, a new technique that requires no input data at all and predicts user item rating data that is competitive in Mean Absolute Error and fairness metric when compared to classic matrix factorization with affluent data, and performs much better than random placement. The authors came to the conclusion that the implemented method outperforms the random recommendation by a wide margin. Even when compared to the standard matrix factorization, the method is competitive in terms of both MAE and fairness. The implications are twofold: 1. The strategy provides an excellent solution to the cold-start problem. 2. Matrix factorization approaches are far from adequate.

The authors [23] suggest a novel technique known as PP-Rec: News Recommendation with Personalized User Interest and time-aware news popularity. The ranking score in this method for recommending a candidate news to a target user is a combination of a personalized matching score and a news popularity score. The former is used to track individual user interest in news. The latter is utilized to calculate the time-aware popularity of candidate news, which is projected using a unified framework based on news substance, recency, and real-time CTR. Furthermore, they suggest a popularity-aware user encoder to remove popularity bias in user behaviors for accurate interest modelling. Experiments on two real-world datasets show that our strategy can significantly enhance news suggestion accuracy and diversity.

The authors of this study [24] offered a study on Facing the cold start problem in recommender systems, i.e. the cold start problem. The authors presented a model in which well-known classification algorithms, together with similarity approaches and prediction mechanisms, give the means for retrieving recommendations. The suggested method includes categorization methods in a pure CF system, and the incorporation of demographic data aids in the identification of additional users who exhibit similar behaviour. Through a vast number of experiments, these experiments demonstrate the performance of the proposed system. In addition, they used the well-known dataset provided by the GroupLens research group. They concluded that the proposed method has advantages by offering satisfactory numerical results in many experimental circumstances.

The author has proposed a brief study on Resolving Cold Start Problem in Recommendation System Using

Demographic Approach. This study [25] focuses on system recommendations for new users. For such a system, a demographic trend was followed by identifying similarities between old and new users. The suggested work is based on movie recommendation and was created with the MovieLens Dataset. This research is based on the similarity factor, which solves the shortcomings of standard Collaborative Filtering and Content-Based Filtering algorithms. When a new or very cold user enters the system, the collaborative filtering strategy fails. The study concluded that content-based filtering failed in terms of performance, while the proposed solution solved the Cold Start Problem. When new goods are introduced to the system, previous users receive a suggestion. Scalability, the major obstacle of Cold Start Problem, has also been addressed in this context for commercial sites.

The authors of this paper [26] conducted research on hybrid recommendation system to solve cold start problem. This page summarizes numerous methodologies such as hybridization procedures, data gathering approaches, the most widely used feasible cold start solutions, regularly used datasets, algorithms, assessment methods, and so on. This paper investigates how current hybrid methods to tackling the cold start problem might assist researchers establish a direction for solving the cold start challenge.

Another cold start research has been proposed, in which the authors [27] address the Extreme Cold-Start Problem in Group Recommendation. Because many cold start solutions have failed to function, the authors created a group recommendation model for EXtreme cold-start in group REcommendation (EXTRE) that is appropriate for the extreme cold start situation. EXTRE's main idea is to leverage the limit theory of graph convolutional neural networks to construct implicit connections between groups and objects, and this derivation does not require explicit interaction data, making it suited for cold start scenarios.

Unlike the growing interest in this topic, the cold-start problem is often regarded as more difficult and intractable. Few researchers [28-30] have thrown some illuminating light on the matter, and the most of them require some form of data input as a starting point to tackle the problem. Since many academics have published several papers on cold start challenges, there is little to no literature accessible where a machine learning technique is applied to tackle the cold start problem. This prompts us to conduct an experimental examination of the machine learning algorithm on the news data and determine how accurate and practical the model is for recommending news to a new user.

3. RECOMMENDATION SYSTEM: A GENERIC APPROACH

There are two sorts of entities in a typical recommender system. The first entity is a user, and the second is an item, and each of these entities is associated with a plethora of rich metadata as shown in figure (Figure 1). We may provide demographic data as well as activity data and pertinent information to the user [31]. As a result, the system must consider all of the data and make better recommendations as a result.

One of the most important types of data is events, which include a variety of information that is obtained anytime a user interacts with the system. These preferences might include page visits, watch history, app interaction, purchases, buying

items, online orders, reviewers, like dislikes, and follows, among other things [32]. The following (Figure 2) is a graphical representation of the events that may be utilized to obtain information about a certain user or an item.

```
Users
{
  "used_id" : 165,
  "user_name" : "recommender system",
  "user_location" : "Loc 123",
  "user_dob" : 20221205,
  "user_profession" : "Std",
  "user_gender" : "Male",
  ...
}

Items
{
  "item_id" : 516,
  "item_name" : "Fruits",
  "item_details" : [color: "yellow", shape: "round"],
  "item_likes" : 216563,
  "item_matrix" : "High",
  "item_value" : "Good",
  ...
}
```

Figure 1. Metadata of users and items entities in a recommender system

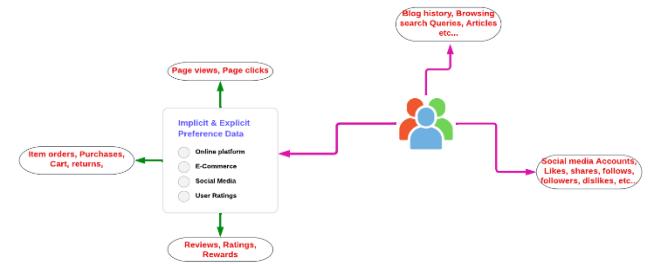


Figure 2. Various events of a user

When a new member joins a social platform, whether online or offline. Some suggestions will be made right away. We can, for example, advocate particular articles or purchase things. We need certain information from the user ahead of time, such as his or her geographical location, gender, age, and so on, so that a system can suggest what a person is interested in. Each event is associated with a set of contexts such as the location, kind of event, ID, device on which it is operated, IP/MAC addressing, time when it is utilized, and most importantly, the type of material. As illustrated in Figure 3, this is one of the most crucial aspects where a system might make suggestions based on the context. The context location in Figure 3 is a DMS latitude location of Jammu and Kashmir [32-34]. In this study, we attempted to develop a framework in which location (IP addressing) plays a significant role. The framework was afterwards validated and tested on various datasets using a conventional machine learning technique, as will be shown in the following sections.

CONTEXT

```
{
  "used_id" : 165,
  "user_name" : "CONTEXT",
  "Context_location" : "34°5'1.1616' N",
  "Context_type" : "Media",
  "Context_time" : 124566000,
  "Context_histroy" : "Based on searches",
  "Context_platform" : "Linux",
  "Context_browser" : "Safari",
  "Context_date" : "20221250",
  "Context_addressing" : "IP addressing",
  ...
}
```

Figure 3. Information about context in a recommender system

The recommendation system filters the material internally and suggests various features to each user. A recommender system suggests several ways to a user. These include dependent on the user's current geographical location. Assume we don't have any information about a user other than their location [35]. If a user is now residing in Kashmir - "Heaven on Earth," the recommender system will first offer tourism destinations to the user based on their current location. Localized suggestion is the name given to this form of recommendation [36].

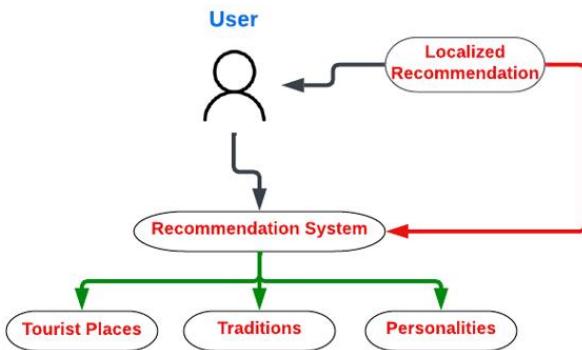


Figure 4. An example of localized recommendation system

In the above figure (Figure 4), a new user whose location is only known is recommended with prominent tourism destinations, traditions, and famous persons. These choices are made because these are the areas that are most frequently utilized when a location (Kashmir) is used [37]. Other forms of recommendations include content-based, collaborative-based, hybrid-based, switching-based, and mixed hybridization, among others. All of these suggestion algorithms are based on the user's history, likes, dislikes, ratings, and so on. These tactics are critical in resolving a cold-start problem [38].

3.1 Current challenges

Many recommendation systems rely on past data about the user or the object to provide recommendations regarding products or services. We are now dealing with two sorts of cold start problems: item-based cold start problems and user-based cold start problems. It is quite difficult for the system to offer items to a new user whose information is not present in the system in a user-based cold start problem [39]. That is, the system is unaware of the new user's interests, hobbies, or any other aspects of his or her daily life. Furthermore, when a new product is introduced and no data is provided since it has not yet been rated [40]. It is also difficult to endorse a product because we have got no information. This kind of cold start problem is known as an item-based cold start problem.

4. EXPERIMENTAL SETUP

In this article, we attempt to address these issues by assuming that in the instance of a user, we have a single set of information about the user, and that information is the user's location. Because, every time a person connects to the internet, he or she must create an account (Gmail, Yahoo, Rediffmail, etc.) that records the user's location while the user is using the internet.

We built the iterative Dichotomizer 3 (ID3) method [41,

[42], which constructs a decision tree on the news dataset and then predicts the values for the test data. The decision tree was chosen for this study because it is still regarded as one of the best and most fundamental classical algorithms. It is highly effective and can be trained on small datasets. The news dataset was obtained from Kaggle and includes BBC news [43] with eight parameters. News ID, News text, Keywords, date, Location, Shares count, followers, and so on are among the parameters. The news is becoming more popular as a result of its rating. The first raw dataset has around 20K records. The dataset was divided into (80-20)% ratios in training and test data. The overall ID3 technique is depicted in the figure below (Figure 5):

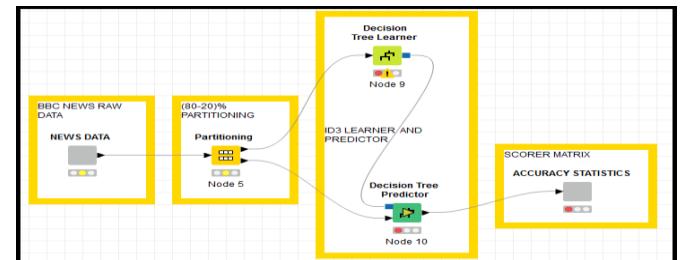


Figure 5. Implemented methodology workflow

Following the construction of the decision tree based on the given dataset, it was compared to the test set and additional statistical parameters were derived using the scorer matrix, as described above. During the partitioning step, 80% of the dataset was submitted to the decision tree learner, with the remaining 20% going through the decision tree predictor. The scorer node calculated all of the accuracy data, which included precision, recall, and other factors. This experiment was carried out using KNIME [44] (an online data manipulation platform), and all raw news data pre-processing was done in Python and its libraries.

4.1 News data aggregation and preprocessing

A user is inundated with information from many sources, and it is not unusual for a user to go between news portals or read news from portals that combine news pieces from several sources (Yahoo! News). With so much information available, it might be tough to choose the news item that a user would enjoy. As a result, people cease or reduce their consumption of news. Because in our scenario, we just know one piece of information about the user, which is their location. As a result, it is extremely difficult for a recommender system to recommend the news to a naive user [45].

Figure 6. A sample portion of news dataset

So, to solve this problem, we adopted a two-step procedure in which we first tried to examine the feedback of each news item and their rating, such as how many times the news has been shared, liked, followed, and so on. Furthermore, we

attempted to organize the news into categories such as political, environmental, sports, health, humour, artistic, traditional, educational, and wedding news, among others [46]. Following that, we also attempted to determine which news items are now highly indexed and rated. Below (Figure 6) is the screenshot of the portion of the dataset on which the experiment was carried out.

4.1.1 Step 1: Pre-processing of news dataset

The above figure consists of (Figure 6) news dataset has several fields such as news ID, News, category (kind of news), date, news keywords, number of shares, location, and so forth. We examined the location-based news (network IDs) and the type of news that was shared, and found that it was relatively evenly distributed, implying that there is no bias in our dataset. The boxplot of news shared with different groups is shown in the picture below (Figure 7).

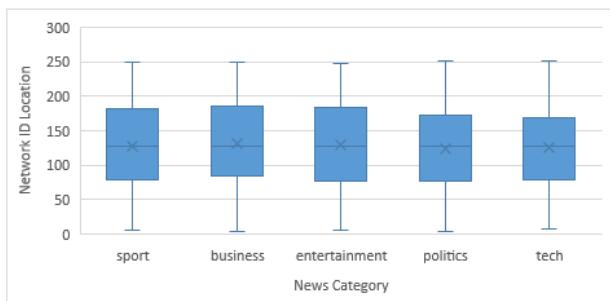


Figure 7. Boxplot of news item shared with different locations

Also, on one specific day, we discovered that business news was shared the most, and it was computed that business news accounted for about 90% of all news shared [47]. As a result, we will mainly choose to share the same type of news with the naive user. But first, we must determine the location of our naive user. The graph below (Figure 8) depicts the most popular news categories on a given day.

The line and the bars in the figure below represents the relative portion of each factor to the total and it shows the most significant factor in the data.

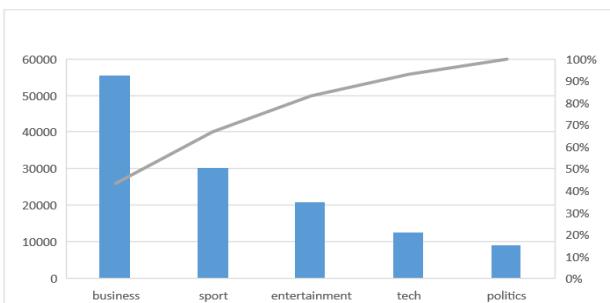


Figure 8. Popularity among news categories based on number of shares

The final step in step 1 is to determine the most popular news based on the shared location, and based on that, we will receive the news that is highly rated among the many categories and areas. Figure 9 depicts the most popular news that was shared by location and category.

In step 1 of the experiment, it was discovered that business news was the most shared of the categories, and this was also

true when we looked at news popularity based on location and news shares. So, based on our dataset and on a certain day, we may infer in step 1 that business news was shared the most among the other categories [48-51].

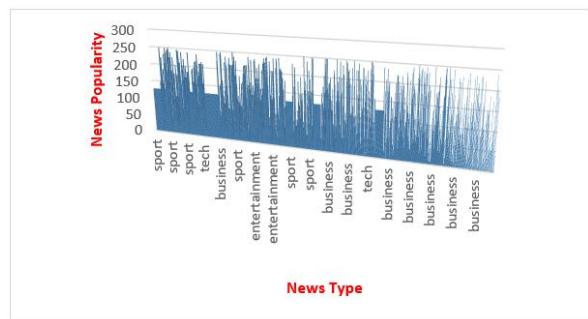


Figure 9. Popularity among news categories based on location and news shares

4.1.2 User based news recommendation

New users and items are added to every recommender system. However, these fresh data records have no prior history, which makes the suggestions very difficult. This is known as the cold start problem, and it is something that every recommender system creator in the business faces. Some products use random or popular items to recommend to new users, while others utilize content-based recommendation in specific settings, such as news recommendation, to overcome the cold-start problem for new items. Personalization is not possible in circumstances of cold-start, when a user who is new to the site or is not logged in comes. Thus, in this step our main objective is to suggest the news to the new user whose location is known only.

Suppose we have many naïve users who wanted to join the internet and we only know the locations of the users from where they are accessing the internet. Below (Table 1) table shows the sample screenshot of the new users with available information. Some of the parameters in the table are displayed as Null since we don't know anything about the new user other than their location and will be suggesting news or other items based on that.

Table 1. User information data

Users	Name	Gender	Location	Browsing history
User_1	NA	NULL	62.207.118.91	NULL
User_2	NA	NULL	62.105.64.169	NULL
User_3	NA	NULL	12.21.198.25	NULL
User_4	NA	NULL	214.113.97.200	NULL
User_5	NA	NULL	4.8.221.188	NULL
User_6	NA	NULL	31.195.181.155	NULL
User_7	NA	NULL	193.86.21.181	NULL
User_8	NA	NULL	225.124.72.127	NULL
User_9	NA	NULL	127.79.90.128	NULL
User_10	NA	NULL	127.17.125.161	NULL
User_11	NA	NULL	127.181.218.85	NULL
User_12	NA	NULL	127.140.17.238	NULL
User_13	NA	NULL	127.181.65.153	NULL
User_14	NA	NULL	127.200.115.157	NULL
User_15	NA	NULL	127.200.115.158	NULL

Therefore, in order to propose a news item to a new user, we first verify the user's location and then compare it to the training set of our data. Assume a user with network ID

127.200.115.157 joins the network and the system wishes to recommend some news to this user. This network ID is compared to the nearest network ID in the training dataset, and the algorithm then predicts some news that is currently trending in this area. As demonstrated in exhibit 9, the network IDs ranging from 127.200.98.56 to 127.212.112.56 had the most business and sports news exchanged. As a result, the system offers the same type of news to this new user by default.

5. EXPERIMENTAL EVALUATION RESULTS: PERFORMANCE

The primary goal of this study is to provide a comprehensive overview of the solution to the Cold-Start issue using a traditional machine learning technique. In this study, we used cutting-edge technology on the BBC news dataset to identify the algorithm with the greatest overall performance and accuracy. In this research, a technique to the step-by-step implementation of the decision tree on the News data is suggested. We have around 16K records of the news as the training set and 4K records are used in testing purposes. After the implementation of the training data is completed we tried to analyze the results on the test dataset. The performance evaluation of ID3 is displayed in the table (Table 2) below:

Table 2. Accuracy statistics

News specifications	Decision tree (ID3)
Training Data (News)	16005
Test Data (Users)	4015
Correctly Classified	3135
Wrongly Classified	880
Accuracy Measure	78.09%
Precision	86.22

According to the tabular representation, we accurately suggested the news to 3135 users out of 4015 users, indicating an accuracy level of roughly 78%. This indicates that the news provided to the new user was anticipated properly about 78%, giving the algorithm a degree of positiveness when evaluated on the testing data with a precision value of 86.22. These reveal some promising findings, and news via search can display fluctuations and reach the level of ranking instability, causing it to stabilize a little later, but access from social networks has an influence on view counts and ranking score. Because we just predicted the news depending on the location in order to deal with the cold start issue and we attained some promising results.

6. DISCUSSION

Numerous scholars from throughout the world have conducted studies on news suggestion, with diverse outcomes. Some recent contributions include classical, ensembled, and ANN approach-based models, among others. Many of these developed algorithms use all of the user's information to predict the user's news. However, we have merely obtained the user's location and applied the ID3 algorithm based on that. We picked the classical technique over the neural network because the decision tree retains its credibility. We analyzed the news and, based on the ranking score, such as the number of shares and the location, we identified the most significant

news on a given day and recommended it to the user. As a result, a mathematical foundation for its Python implementation on Google Colab was devised and built. Following implementation, ID3's performance was roughly 78%. The final accuracy statistics of the ID3 on the BBC news dataset yield remarkable results with a high level of correctness.

However, the influence of ID3 on other news datasets was not investigated in this study. The findings in this work are inevitably constrained, and they exclude several other research trends. As a result, it is strongly advised that the same implementation be tested on multiple datasets.

7. CONCLUSION AND FUTURE WORK

The ID3 approach was suggested utilizing BBC news data, which comprises 8 explanatory variables, including News ID, News text, Keywords, date, Location, Shares count, followers, and so on. One of the best fit models developed using (80-20%) training and test ratios achieved a classification accuracy of roughly 78%. Our solution is an effective tool for the cold-start problem since it outperforms the guideline by a significant margin depending on the locale. According to the results, our technique is competitive in terms of both accuracy and precision. The main advantage of building this framework is just to simply suggest the items particularly the news to the new user whenever a person connects to the internet.

Nonetheless, we only employed one BBC news dataset and a small number of users; what about utilizing numerous news datasets, and how does the rise in the number of online users influence system performance? Is the active node approach adaptable to individual system requirements? Is the incorporation of the active node approach into the semantic structure enhancing levels of novelty for node recommendation and/or coverage and precision for batch recommendations? All of these questions must still be answered.

REFERENCES

- [1] Gulla, J.A., Zhang, I., Liu, P., Özgöbek, Ö, Su, X. (2017). The addressa dataset for news recommendation. WI '17: Proceedings of the International Conference on Web Intelligence, pp. 1042-1048. <https://doi.org/10.1145/3106426.3109436>
- [2] Zhang, Q., Li, J., Jia, Q., Wang, C., Zhu, J., Wang, Z., He, X. (2021). UNBERT: User-news matching BERT for news recommendation. In IJCAI, pp. 3356-3362.
- [3] IJntema, W., Goossen, F., Frasincar, F., Hogenboom, F. (2010). Ontology-based news recommendation. In Proceedings of the 2010 EDBT/ICDT Workshops, pp. 1-6. <https://doi.org/10.1145/1754239.1754257>
- [4] Li, L., Li, T. (2013). News recommendation via hypergraph learning: Encapsulation of user behavior and news content. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 305-314. <https://doi.org/10.1145/2433396.2433436>
- [5] Cantador, I., Castells, P., Bellogín, A. (2011). An enhanced semantic layer for hybrid recommender systems: Application to news recommendation. International Journal on Semantic Web and Information

- Systems (IJSWIS), 7(1): 44-78. <https://doi.org/10.4018/jswis.2011010103>
- [6] Yeung, K.F., Yang, Y. (2010). A proactive personalized mobile news recommendation system. In 2010 Developments in E-systems Engineering, pp. 207-212. <https://doi.org/10.1109/DeSE.2010.40>
- [7] Kaul, S., Zaman, M., Fayaz, S.A., Butt, M.A. (2023). Performance stagnation of meteorological data of kashmir. In: Gupta, D., Khanna, A., Bhattacharyya, S., Hassani, A.E., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. Lecture Notes in Networks and Systems. https://doi.org/10.1007/978-981-19-2535-1_63
- [8] Fayaz, S.A., Jahangeer Sidiq, S., Zaman, M., Butt, M.A. (2022). Machine Learning: An Introduction to Reinforcement Learning. Machine Learning and Data Science: Fundamentals and Applications, 1-22. <https://doi.org/10.1002/9781119776499.ch1>
- [9] Fayaz, S.A., Zaman, M., Kaul, S., Butt, M.A. (2022). How M5 model trees (M5-MT) on continuous data are used in rainfall prediction: An experimental evaluation. *Revue d'Intelligence Artificielle*, 36(3): 409-415. <https://doi.org/10.18280/ria.360308>
- [10] Kaul, S., Fayaz, S.A., Majid, Z., Butt, M.A. (2022). Is decision tree obsolete in its original form? A burning debate. *Revue d'Intelligence Artificielle*, 36(1): 105-113. <https://doi.org/10.18280/ria.360112>
- [11] Lam, X.N., Vu, T., Le, T.D., Duong, A.D. (2008). Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, pp. 208-211. <https://doi.org/10.1145/1352793.1352837>
- [12] Rehman, A., Butt, M. A., Zaman, M. (2022). Liver lesion segmentation using deep learning models. *Acadlore Transactions on AI and Machine Learning*, 1(1): 61-67. <https://doi.org/10.56578/ataiml010108>
- [13] Tahmasebi, F., Meghdadi, M., Ahmadian, S., Valiollahi, K. (2021). A hybrid recommendation system based on profile expansion technique to alleviate cold start problem. *Multimedia Tools and Applications*, 80(2): 2339-2354. <https://doi.org/10.1007/s11042-020-09768-8>
- [14] Gope, J., Jain, S.K. (2017). A survey on solving cold start problem in recommender systems. In 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 133-138. <https://doi.org/10.1109/CCAA.2017.8229786>
- [15] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Numerical and experimental investigation of meteorological data using adaptive linear M5 model tree for the prediction of rainfall. *Review of Computer Engineering Research*, 9(1): 1-12. <https://doi.org/10.18488/76.v9i1.2961>
- [16] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall. *International Journal of Advanced Technology and Engineering Exploration*, 9(89): 509-522. <http://dx.doi.org/10.19101/IJATEE.2021.874647>
- [17] Liao, M., Sundar, S.S. (2022). When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *Journal of Advertising*, 51(2): 256-267.
- https://doi.org/10.1080/00913367.2021.1887013
- [18] Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Mahboob Alam, T., Luo, S. (2021). A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning (iJET)*, 16(3): 274-306.
- [19] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). A super ensembled and traditional models for the prediction of rainfall: An experimental evaluation of DT versus DDT versus RF. In: Sharma, H., Shrivastava, V., Kumari Bharti, K., Wang, L. (eds) *Communication and Intelligent Systems*. Lecture Notes in Networks and Systems. https://doi.org/10.1007/978-981-19-2130-8_48
- [20] Okura, S., Tagami, Y., Ono, S., Tajima, A. (2017). Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1933-1942. <https://doi.org/10.1145/3097983.3098108>
- [21] Embarak, O.H. (2011). A method for solving the cold start problem in recommendation systems. *2011 International Conference on Innovations in Information Technology*, pp. 238-243. <https://doi.org/10.1109/INNOVATIONS.2011.5893824>
- [22] Wang, H. (2021). ZeroMat: Solving cold-start problem of recommender system with no input data. In *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 102-105. <https://doi.org/10.1109/ICISCAE52414.2021.9590668>
- [23] Qi, T., Wu, F., Wu, C., Huang, Y. (2021). PP-Rec: News recommendation with personalized user interest and time-aware news popularity. *arXiv preprint arXiv:2106.01300*.
- [24] Likha, B., Kolomvatsos, K., Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Syst. Appl.*, 41(4): 2065-2073. <https://doi.org/10.1016/j.eswa.2013.09.005>
- [25] Pandey, A.K., Rajpoot, D.S. (2016). Resolving cold start problem in recommendation system using demographic approach. In *2016 International Conference on Signal Processing and Communication (ICSC)*, pp. 213-218. <https://doi.org/10.1109/ICSPCom.2016.7980578>
- [26] Rahman, M., Shama, I.A., Rahman, S., Nabil, R. (2022). Hybrid recommendation system to solve cold start problem. *Journal of Theoretical and Applied Information Technology*, 100(11): 3562-3578.
- [27] Guo, L., Tao, Y., Gao, M., Yu, J., Li, W. (2022). Addressing the extreme cold-start problem in group recommendation. *arXiv preprint arXiv:2210.09672*.
- [28] Fayaz, A., Zaman, M., Kaul, S., Butt, M.A. (2022). Is deep learning on tabular data enough? An assessment. *International Journal of Advanced Computer Science and Applications*.
- [29] Khojamli, H., Razmara, J. (2021). Survey of similarity functions on neighborhood-based collaborative filtering. *Expert Systems with Applications*, 185: 115482. <https://doi.org/10.1016/j.eswa.2021.115482>
- [30] Koren, Y., Rendle, S., Bell, R. (2022). Advances in collaborative filtering. *Recommender systems handbook*, 91-142. https://doi.org/10.1007/978-1-0716-2197-4_3
- [31] Feng, J., Xia, Z., Feng, X., Peng, J. (2021). RBPR: A hybrid model for the new user cold start problem in recommender systems. *Knowledge-Based Systems*, 214: 106732. <https://doi.org/10.1016/j.knosys.2020.106732>

- [32] Ashraf, M., Salal, Y.K., Abdullaev, S.M., Zaman, M., Bhut, M.A. (2022). Introduction of Feature Selection and Leading-Edge Technologies Viz. TENSORFLOW, PYTORCH, and KERAS: An Empirical Study to Improve Prediction Accuracy of Cardiovascular Disease. In International Conference on Innovative Computing and Communications, pp. 19-31. https://doi.org/10.1007/978-981-16-3071-2_2
- [33] Mohd, R., Butt, M.A., Baba, M.Z. (2020). GWLM-NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction. Data Technologies and Applications, 54(1): 85-102. <https://doi.org/10.1108/DTA-08-2019-0130>
- [34] Fayaz, S.A., Zaman, M., Butt, M.A. (2021). To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: An empirical discourse of geographical data mining. Procedia Computer Science, 184: 935-940. <https://doi.org/10.1016/j.procs.2021.03.116>
- [35] Fayaz, S.A., Zaman, M., Butt, M.A. (2021). An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data. International Journal of Advanced Technology and Engineering Exploration, 8(84): 1424-1440. <http://dx.doi.org/10.19101/IJATEE.2021.874586>
- [36] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Performance evaluation of GINI index and information gain criteria on geographical data: An empirical study based on JAVA and Python. In International Conference on Innovative Computing and Communications, pp. 249-265. https://doi.org/10.1007/978-981-16-3071-2_22
- [37] Zihayat, M., Ayanso, A., Zhao, X., Davoudi, H., An, A. (2019). A utility-based news recommendation system. Decision Support Systems, 117: 14-27. <https://doi.org/10.1016/j.dss.2018.12.001>
- [38] Kanwal, S., Malik, M.K., Nawaz, Z., Mehmood, K. (2022). Urdu wikification and its application in urdu news recommendation system. IEEE Access, 10: 103655-103668. <https://doi.org/10.1109/ACCESS.2022.3208666>
- [39] Zhong, M., Ding, R. (2022). Design of a personalized recommendation system for learning resources based on collaborative filtering. International Journal of Circuits, Systems and Signal Processing, 16: 122-131. <https://doi.org/10.46300/9106.2022.16.16>
- [40] Yang, T., Zhang, J., Wang, L., Zhang, J. (2022). A novel customer-oriented recommendation system for paid knowledge products. Journal of Systems Science and Systems Engineering, 31(5): 515-533. <https://doi.org/10.1007/s11518-022-5540-x>
- [41] Butt, E.M.A., Quadri, S.M.K., Zaman, E.M. (2012). Star schema implementation for automation of examination records. In Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS).
- [42] Butt, M.A., Zaman, M. (2013). Assessment model based data warehouse: A qualitative approach. International Journal of Computer Applications, 62(10): 22-24.
- [43] Taj, S., Shaikh, B.B., Meghji, A.F. (2019). Sentiment analysis of news articles: A lexicon based approach. In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-5. <https://doi.org/10.1109/ICOMET.2019.8673428>
- [44] Altaf, I., Butt, M.A., Zaman, M. (2022). ETL for disease indicators using brute force rule-based NLP algorithm and metadata exploration. International Journal of Advanced Technology and Engineering Exploration, 9(90): 644-652. <https://doi.org/10.19101/IJATEE.2021.875069>
- [45] Altaf, I., Butt, M.A., Zaman, M. (2022). Machine Learning Techniques on Disease Detection and Prediction Using the Hepatic and Lipid Profile Panel Data. In Congress on Intelligent Systems, pp. 189-203. https://doi.org/10.1007/978-981-16-9113-3_15
- [46] Camacho, L.A.G., Alves-Souza, S.N. (2018). Social network data to alleviate cold-start in recommender system: A systematic review. Information Processing & Management, 54(4): 529-544. <https://doi.org/10.1016/j.ipm.2018.03.004>
- [47] Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., Cai, H. (2019). Addressing the item cold-start problem by attribute-driven active learning. IEEE Transactions on Knowledge and Data Engineering, 32(4): 631-644. <https://doi.org/10.1109/TKDE.2019.2891530>
- [48] Altaf, I., Butt, M.A., Zaman, M. (2021). A pragmatic comparison of supervised machine learning classifiers for disease diagnosis. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1515-1520. <https://doi.org/10.1109/ICIRCA51532.2021.9544582>
- [49] Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., Foster, E. (2019). Recommender system based on pairwise association rules. Expert Systems with Applications, 115: 535-542. <https://doi.org/10.1016/j.eswa.2018.07.077>
- [50] Hasan, I., Rizvi, S.A.M., Zaman, M., Bakshi, W.J., Fayaz, S.A. (2022). A scalable framework to analyze data from heterogeneous sources at different levels of granularity. Information Dynamics and Applications. 1(1): 26-34. <https://doi.org/10.56578/ida010104>
- [51] Fayaz, S.A., Kaul, S., Zaman, M., Butt, M.A. (2022). An adaptive gradient boosting model for the prediction of rainfall using ID3 as a base estimator. Revue d'Intelligence Artificielle, 36(2): 241-250. <https://doi.org/10.18280/ria.360208>