Engaging Minds: Boosting Brand Awareness with AI-Powered Content Recommendations

Final Capstone Project Presented to the Faculty of Science and Technology at IE University in Partial Fulfillment of the Requirements for the Degree of Master in Business Analytics and Big Data (SAMBD)

Prepared by Group 2

Naif Almahdawi
Elias Nicholas Markopoulos
Mishal Albuainain
Hamad Aldemikhi
Rawan Majid Alshahwan
Khalid Al Shigager

Supervised by
Prof. Guillermo de Haro Rodríguez, Vice Dean of Graduate Programs at IE University
Prof. Federico Castanedo Sotela, Adjunct Professor at IE University

AI Disclosure: No content generated by AI technologies has been used in this report. Generative AI was used to support code development and debugging.

I hereby certify that this report and the accompanying presentation are our original work in their entirety unless where indicated and referenced.

Github Repository

IE University
December, 2024

Index

Introduction

Our capstone project, "Engaging Minds: Enhancing Brand Awareness through AI-Driven Content Recommendations," aims to improve digital engagement to support Aramco's global brand impact. Personalization is essential for developing meaningful connections with audiences in the digital age. This project created an AI and machine-learning-powered recommendation engine to offer real-time, personalized content to Aramco's platforms to engage people and build loyalty. The system provides relevant and tailored user experiences via a natural language processing deep learning model. It also tackles the "cold start" issue, guaranteeing that new users without prior data receive useful and interesting material, drawing a wider audience, and strengthening Aramco's image as an innovative and user-centric energy leader.

Owned channels are crucial. AramcoLife, Elements Magazine, and Raceteq help communicate Aramco's ideas, accomplishments, and energy leadership. Aramco can create narratives that resonate with its target audience on owned platforms, unlike paid or third-party channels. Effective content on multiple platforms boosts user engagement, brand awareness, and trust. This recommendation engine helps Aramco optimize its owned media strategy by delivering content to the right audience at the right time, increasing user engagement and brand loyalty.

The project used the Microsoft News Dataset (MIND), which contains over one million users and 160,000 articles, to analyze user behavior, identify patterns, and create a real-time, personalized news recommendation engine.

Objective & Problem Definition

Notable parallels can be drawn between the objective of our project and its application to Aramco's business strategy and IKEA's retail strategy. IKEA designs its stores with a clear goal: to keep customers engaged and in-store for as long as possible. Everything is designed to make customers stay longer and interact with more of the brand's products, from the winding aisles that encourage exploration to the carefully placed displays that spark curiosity[1].

Our project shares a similar goal but in a digital environment. Just as IKEA aims to extend a customer's time in-store, we aim to keep users engaged on corporate websites and publications longer. In both cases, it is about creating a journey tailored to the individual—through a physical store layout or personalized content recommendations. For IKEA, the benefit is clear: the longer customers stay, the more likely they are to discover something they want to buy. For a website, it is about holding users' attention, encouraging them to explore more content, and deepening their connection to the brand. This is especially valuable for corporate websites like Aramco's, where the goal is to inform and build awareness, familiarity, and preference for the brand.

Our team set out to create a model that delivers this personalized experience online. By understanding what users are most likely to engage with, we can guide them through a curated digital journey that keeps them exploring content that matters to them, just as IKEA stores guide shoppers through its products.

This project aims to show how the principles that make IKEA's approach effective in retail can be applied to digital content. With personalization as our focus, we're creating a recommendation system to keep users engaged and connected to a brand for longer.

In today's fast-paced, time-poor world, people expect brands to know and cater to their preferences. This is even more apparent when it comes to consuming content from publications and websites. Personalization of digital content and services has been around since the birth of the Internet, but it has accelerated through the mass adoption of social media.

Let us take the growth of social media, for example. Social media services rely on personalization to serve content to users which they know will appeal to them. Thus keeping them engaged and on their platform for longer. This is important for social media platforms as it allows them to sell the user's time to advertisers.

Personalization for brand publications can work in a similar way. Personalized news recommendations can transform how we see, read, and consume information. It saves time and improves the user's experience by customizing content according to their interests.

Consider a scenario where a sports fan visits a website and is presented with a personalized feed of items relevant to their interests. They are much more likely to engage with this content and consume more similar content. For a publication, this is critical for the same reasons as a social media website. The longer they can keep a user on the website, the more advertising space they can sell to advertisers.

For Aramco, it is a similar strategy, but rather than selling advertising space, the company's content is the advertising, helping to communicate crucial ideas and sentiment about the company.

However, developing an effective news recommendation engine tailored to each user is difficult, as user preferences vary. Compared to social media platforms, websites may not know the user they are serving content to as they do not necessarily need to have a user profile and sign in to the website. Creating a news recommendation system to predict user interest and deliver real-time recommendations with limited information is a complex challenge. It requires a lot of data to train a model on.

Corporate websites are even less likely to require visitors to sign in. While they may be able to have limited information about the user, such as whether they are a new user or a returning user, they are unlikely to know anything about a particular user's preferences. This is known as the cold start issue.

In Aramco's case, the company maintains three major digital publications besides its corporate website: AramcoLife, Elements magazine, and Raceteq. These publications exist to achieve three main goals: increase brand awareness, encourage familiarity with the company, and drive preference for Aramco over its competitors. Ultimately, by implementing a model that predicts which content from Aramco's publications a user will prefer, Aramco can keep users consuming its content for longer and become more familiar with the company.

Our objective is to develop a model that can be applied to any corporate publications to help keep website visitors engaged with the brand's content for longer. We have developed and trained our model using the Microsoft News Dataset (MIND), which contains behavior data of one million users and over 160,000 news articles.

Throughout this journey, we explored several techniques. We combined collaborative filtering, which leverages user-item interactions to recommend similar articles to users with similar preferences, with content-based filtering, which analyzes the content of news articles to recommend similar articles based on topics and similarity. We utilized deep learning models to capture complex patterns in user behavior and their interaction with news, resulting in more accurate and personalized recommendations.

This paper will explore our approach and methodologies to tackle this complex challenge. We will explore the strategies and methods we employed to develop a solution that effectively addresses the problem of personalized recommendations.

Data sources and EDA

To begin to solve this challenge, we began with the large-scale Microsoft News Dataset (MIND), which contains one million users and more than 160k English news articles, each of which has rich information about the users who had at least five news clicks during a span of six weeks from October 12 to November 22, 2019. This data was available in six CSV files containing three CSV files for each training, testing, and validation for user behaviors and three additional CSV files for news articles.

As the datasets provided valuable insights about users and their behaviors, we developed sophisticated recommendation models by analyzing users' click patterns and combining them with insights gathered from article content.

The behavior dataset included [Impression ID, User ID, Time, History, Impressions] features, and the news dataset included [News ID, Category, SubCategory, Title, Abstract, URL, Title Entities, Abstract Entities]. We immediately knew we had to perform essential data preprocessing steps to ensure an efficient machine-learning pipeline and prepare it for analysis. Our preprocessing steps focused on the string data in the news dataset, as most of it included multiple unstructured features. To derive meaningful insights and reduce the computational impact, we extracted the most relevant information while minimizing noise to ensure the quality of our model. Additionally, we removed the URL, Title and Abstract Entities features as they had minimal contribution to the semantic understanding of the text and could potentially introduce noise into the analysis. We correlated each user with news preferences based on click data for the behavior dataset. The history feature provides all news that the user has clicked on before the impression date. The impression feature provides a list of current news on a specific date that shows the user's activity, with 1 for clicked and 0 for not clicked. Therefore, the team was able to understand the importance of extracting each user's clicked and not clicked news and created two additional features called [Clicked News Id, Not Clicked News Id]. We decided to make the recommendation system more holistic and tailored to each user, whether existing or new users, whose history of impressions may not be recorded. To tailor our recommendation system to a broader spectrum of users, we revised our approach, code, and strategy by eliminating the time aspects of our dataset, which helped improve the recommendation system and reduce the computational overhead.

Exploratory Data Analysis is essential to any machine-learning project. It helps us understand the data, identify hidden patterns, and derive insights to build our recommendations model. Since our project is about a news recommendation system, we started our analysis by identifying which news category is more prominent in our datasets.

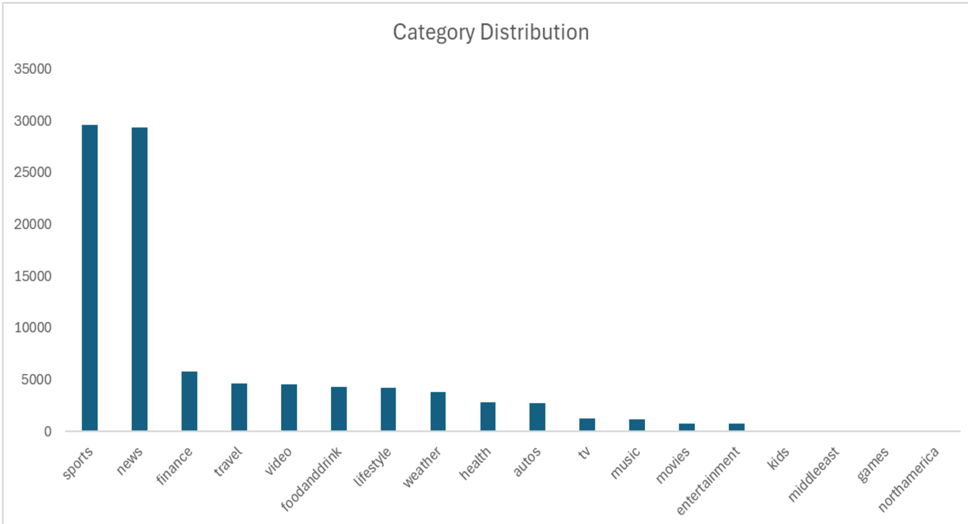


Figure-1: News Category Distribution

Figure 1 illustrates the distribution of news across all categories, showing sports and news dominance over other categories such as finance, travel, video, etc., with a distribution of 29,625 and 29,363, respectively. As Title and Abstract columns are essential to our prediction model, we decided to analyze the length distribution of both columns to evaluate if there is any correlation with user clicks.

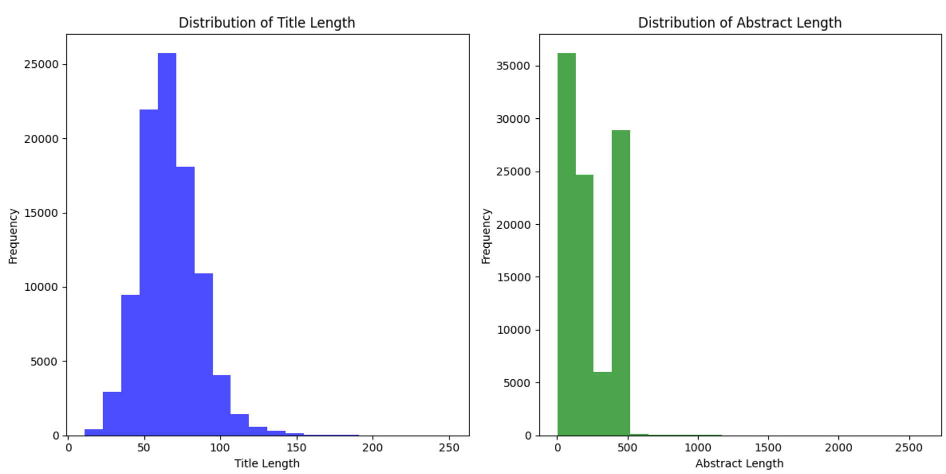


Figure-2: Distribution of Title and Abstract Length

Figure 2 demonstrates the Title length normal distribution where most news titles are between 50-100 characters. The abstract length distribution differs from the Title distribution. Most Abstracts are either 100 or 400 characters, with very few abstracts that reach beyond 1000 characters.

We expanded on these findings in Figure 3 (in the appendix) to evaluate if the title length strongly correlates with the number of clicks. Both charts show a similar trend where the most clicked news is between 50-100 characters.

Next, we analyzed the users' behavior regarding their click patterns. Figure 4 (in the appendix) shows fascinating user profiling behavior, with the vast majority of users clicking on very few articles (0-5 clicks), which suggests that most of the dataset's users are passive viewers who interact with few articles.

User activity is an integral part of our dataset. Therefore, knowing when the users are browsing the website and clicking on their favorite news is demonstrated in Figure 5 (in the appendix). The figure shows that most users are active between 6 AM-1 PM, the rest are distributed across the evening and early morning. Figure 1 shows the news category distribution. But how are these categories correlated with users' clicks? Do the most frequent news categories match the user clicks?

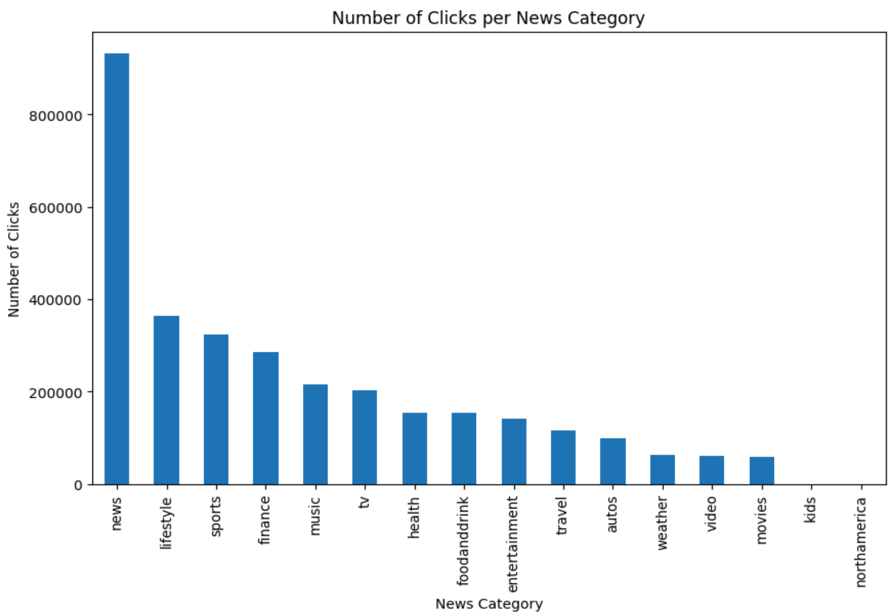


Figure-6: Number of Clicks per News Category

Figure 6 provides an answer to this question, showcasing the news category being more prominent than other categories, with a staggering click of more than 800,000 clicks compared to the second highest clicked category, which is lifestyle, which is shy of 400,000. These details will become a fundamental part of our personalized news recommendations system.

The history column in the behaviors dataset shows the user's clicks before the current impression. The impression column shows the user's interaction with the news displayed, whether they clicked or not. This brings us to the question of whether users' current click behavior is consistent with their historical clicks. Figure 7 illustrates the match rates across all users where two peaks are evident.

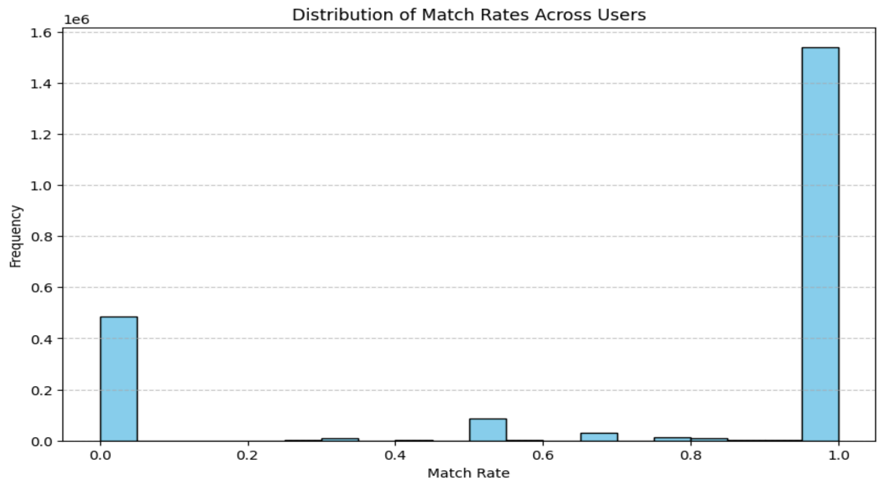


Figure-7: Distribution of Match Rates

Most users like what they previously surfaced with a match rate of 1.0 and the other peak shows the second category of users that like to explore new content rather than their historical preference.

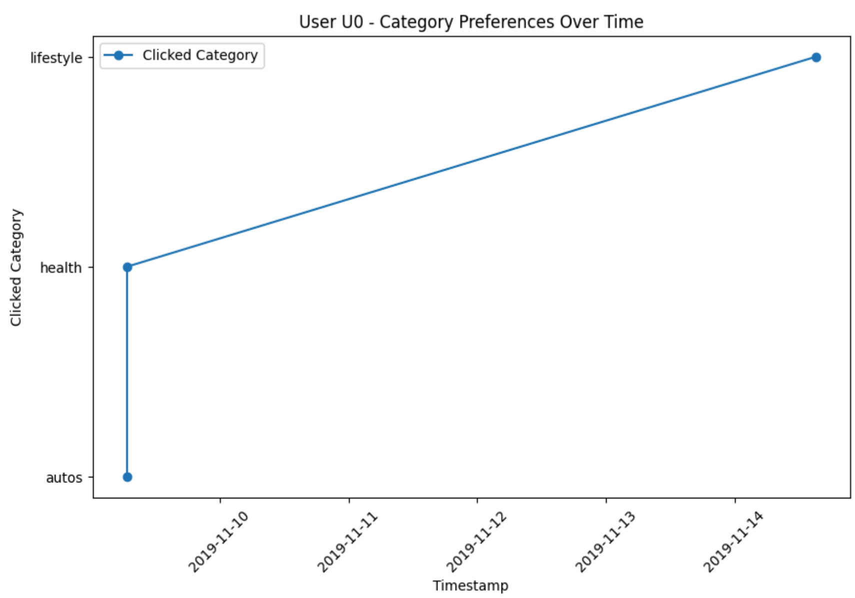


Figure-8: User U0 Category Preference

Moreover, in Figure 8, we analyzed User [U0] clicking behavior and saw his clicking preference. This user clicked on two news articles related to health and autos on 11/9/2019; and then started exploring a different category on 11/14/2019 by clicking on lifestyle news. This dynamic change in behavior will add complexity to our prediction system, as people's preferences can change dynamically. Since we established that the most clicked category is news, followed by lifestyle, we wanted to analyze the distribution of clicks per category.

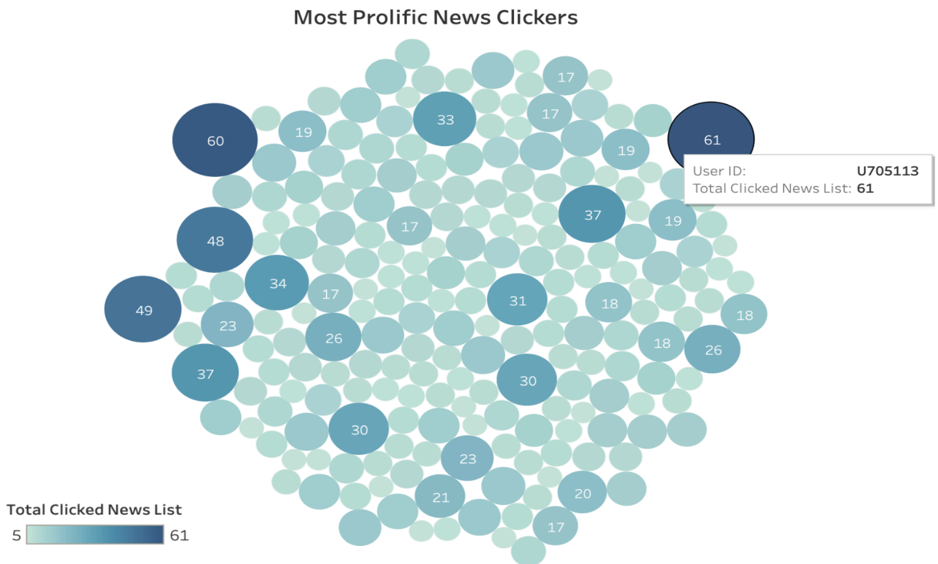


Figure 9 - Sampled click volume distribution

Figure 9 shows that most users are considered low-clickers with 0-5 clicks, followed by moderate clickers with 6-15 clicks, and few with more than 15 clicks. Interestingly, there were no clicks on kids or North America categories, demonstrating that our users are not interested in this type of content. Additionally, we analyzed users' click behavior each day to see whether it was increasing during weekdays or weekends.

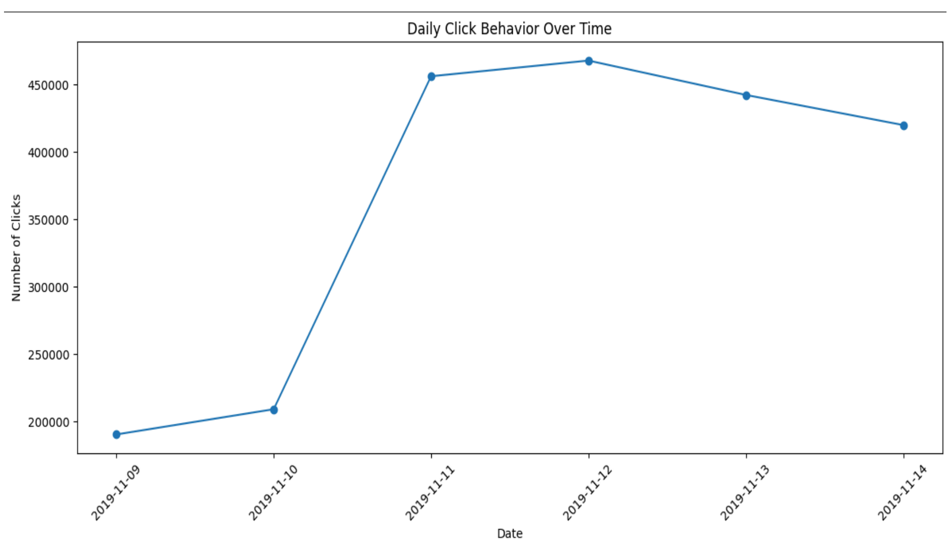


Figure-10: Daily Click Behavior Over Time

Figure 10 shows that most clicks happen during weekdays, and if we combine that with Figure 5, we have the heatmap in Figure 11 that shows the color intensity combining the days and hours of user engagement.

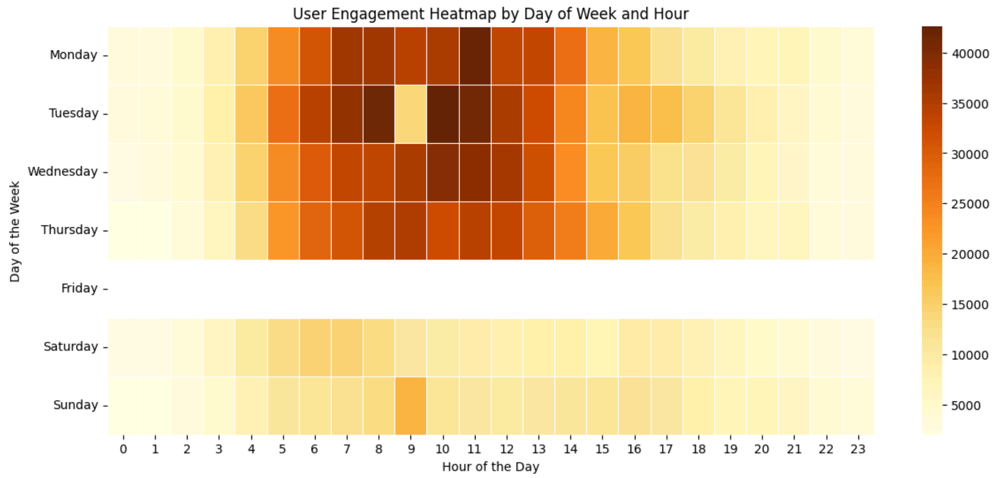


Figure-11: User Engagement Heatmap

Another critical finding was analyzing the news sentiment to assess whether users' engagement correlates with positive, negative, or neutral news.

Figure 12 (in the appendix) shows the distribution of users' clicks across all news, with a clear separation of neutral news being clicked the most compared to positive and negative news. Additionally, we combined and standardized the news IDs by removing the extra characters such as brackets and quotes from our created columns "Clicked News IDs" and "Not-Clicked News IDs," which resulted in a total count of unique news IDs of 27031 in comparison to the original unique number of 72923. Furthermore, we compared this unique news with the historical news clicked by all users, the result shows there were 4723 overlapping news. This analysis helped us identify duplicate records to guide us further in ensuring data consistency across all of our datasets.

All of these analyses were conducted across the training dataset to aid us in moving forward to the validation datasets. After a comprehensive inspection, the validation datasets were similar to the training datasets. Therefore, the same data pre-processing pipeline was implemented to ensure data integrity, consistency, and quality. Moreover, the team analyzed the distribution and consistency of displayed news between the training and validation datasets.

Figure 13 (in the appendix) shows a Venn diagram that compares both datasets, where the overlap illustrates that most news exists between datasets, and the training dataset has more than 13,000 unique news, representing only 13% of all news across both datasets.

Furthermore, we also analyzed how many users will not be part of the training phase. Figure 14 (in the appendix) illustrates that around 15% of the users in the validation datasets are expected to be new to the model, as they were not part of the training phase. Hence, we have built a default user profile to resolve the cold start issue. However, the user interface app includes interaction buttons that will capture the user feedback on the displayed content. Thereby, these feedbacks will serve as an input data during the continuous integration and continuous deployment, to keep the model updated and ready to feed new users their preferred content.

Potential Solutions

We researched several different recommendation models to identify potential solutions to our problem. First, we considered content-based recommendation models, which focus on the properties of news articles. For example, features such as the title, abstract, category, and named entities are extracted and used to recommend other articles with attributes similar to a user's preference. Content-based recommendations are very useful for datasets with rich metadata and are particularly effective for new users for whom we need more historical behavior data. However, content-based recommendations struggle when users have broader preferences, which may change over time, as evidenced by Figure 8. As we build a model we aim to implement for Aramco publications, we must ensure that the model can handle regularly updated user preferences and adapt over time. For this reason, we opted not to use content-based recommendations.

Next, we explored collaborative filtering models. To recommend articles, collaborative filtering relies on user interactions, such as clicks, impressions, and ratings. The key idea is to find patterns between users and news consumption habits. There are benefits to this approach, as using this model type we can discover relationships between users and news articles that are not obvious through content alone, leading to more diverse recommendations than content-based methods. This approach is viable as it relies on large amounts of user data in the MIND dataset. However, this approach would suffer from the cold start issue, particularly for new users.

Other approaches include Deep Learning-based and Reinforcement Learning-based models, which we discounted as a viable option for this project, as they require a lot of computational power, are challenging to interpret, and are complex to deploy in real-world applications.

A hybrid model could be built based on the strengths of content-based and collaborative filtering models. This approach would reduce limitations such as cold start by leveraging content-based filtering for new articles/users and can offer more personalized and diverse recommendations.

Ultimately, we developed a robust hybrid recommendation system combining content-based filtering with collaborative filtering techniques, leveraging advanced preprocessing of article content and user interaction data. Initially, we focused on content-based methods, extracting rich semantic representations of news articles. However, adding user interaction data to the behavior dataset significantly enhanced the personalization capabilities of our system by introducing collaborative filtering elements.

The preprocessing of the behavior dataset was a critical step in integrating user interaction data effectively. The dataset included key columns such as `User ID,` `Displayed News List,` `Clicked News IDs,` and `Not-Clicked News IDs.` These columns required meticulous cleaning and standardization to ensure consistency and usability. By applying regex-based transformations, we converted concatenated and improperly formatted news IDs into clean, comma-separated lists. This step was essential to prepare the data for subsequent processing, ensuring that no malformed entries remained in the interaction columns. We validated this cleaning process by checking for any residual concatenated IDs and aligning data types across columns to maintain compatibility. At the same time, we continued to refine the content-based aspects of our system. Using sophisticated algorithms such as; TF-IDF, LDA, Doc2vec, Word2vec, and ALS to create meaningful features to enhance the content-based filtering performance. However, these features did not yield optimal representation of news articles which led us to explore other potential solutions which is content and collaborative based filtering hybrid approach.

To introduce a collaborative aspect, we utilized the cleaned behavior dataset to identify patterns in user interactions. The `Clicked News IDs` and `Not-Clicked News IDs` columns provided insights into user preferences and non-preferences, enabling us to model user behavior effectively. This data serves as the foundation for collaborative filtering, where recommendations are based on the behavior of similar users (user-based collaborative filtering) or the properties of similar articles. Additionally, the `Displayed News List` column provides contextual data, helping us understand the articles presented to users before their interactions, which adds depth to our analysis.

This approach combines these collaborative filtering elements with content-based features to form an accurate hybrid recommendation model. We also applied K-means clustering to group articles into segments based on one-hot-encoded categorical features like category and subcategory. This clustering step enhances the collaborative filtering component by creating a high-level segmentation of articles, allowing us to identify patterns that may not be evident from individual articles or user interactions alone. By integrating clustering, our system can recommend articles based on content similarity and the broader patterns emerging from grouped features.

This hybrid approach balances personalization and robustness effectively. Content-based filtering addresses the cold start issue for new users. On the other hand, collaborative filtering enriches the recommendations for existing users by leveraging historical interaction data. The combination ensures that our system remains adaptable to different user profiles and provides meaningful recommendations even with varying levels of user engagement. This recommendation system approach demonstrates the power of integrating content-based and collaborative filtering techniques. The preprocessing pipeline ensures that article and user interaction data are clean, structured, and ready for analysis, forming the foundation of our approach. We have created a scalable and highly personalized recommendation engine that effectively serves diverse user needs by leveraging advanced embedding techniques, clustering, and user interaction data. Further enhancements could include deep collaborative filtering methods to capture non-linear relationships and reinforcement learning to optimize for long-term user engagement. This approach performed well in accurate predictions. However, having in mind a model that has the capability of both scalability and immediate inference in real-time deployment, we reached to an optimal solution described in the methodology section as the final deployed method.

Methodology

While exploring different approaches to building a best-in-class news recommendation system, we have tested multiple methodologies for capturing accurate, precise, and meaningful representations of news articles. Information contained in category, sub-category, title, and abstract must be represented in the vector space accurately so that during model training or inference, differentiation will not be a bottleneck for the model or inference function. We explored algorithms, including TF-IDF, word2vec, doc2vec, LDA, K-means clustering, sentiment analysis, emotion analysis, and fine-tuning BERT. The main objective through many exploration and training experiments was to achieve the best possible feature creation that would yield optimal contextual representation. The following sections will shed light on some of these explored feature-creation techniques and, eventually, identify which techniques were the best thus used to create the features that better represent the news articles.

Starting with techniques that were not used due to the limitation of meaningful representation of its purpose, TF-IDF was rejected as it failed to highlight the most important keywords in the text and making this useful for downstream analysis as it does not capture the context of a text rather than the words importances which did not contribute positively to project ultimate goal.

As for word2vec technique, it is usually used to convert words to vectors, but that was not helpful for this news recommendations system use case as this technique will only assign vectors to words rather than being aware of the meaning of a word in a certain text as in English language a word could have many meanings depending on the context of the text. Additionally, out of training scope, all words would be limited for such a technique and would not assign a proper vector for representation.

The doc2vec technique is just an extension of word2vec technique, but instead of word vectorization, it will look-up at the whole document and create a representation vector. However, this approach of document vectorization resulted in representations that are not aware of the context of vector, which is essential information for our use case. In addition, it lacks the ability to vectorize documents out of the training scope; also, it is susceptible to tokenization and stopwords, which limit this technique's ability to be reliable on diverse applications such as news article representation.

Latent Dirichlet Allocation (LDA) is a technique to assign probabilistic representations of different texts within a document e.g., 60% of the news article is politics while 40% is finance. This technique looked promising initially, but after further research, we discovered that it lacks the ability to consider the context and the word order in any text. It also requires enormous training and adoption in any specific domain of application. Furthermore, it is sensitive to tokenization and stopwords similar to doc2vec technique.

Now, exploring one of the used techniques, k-means, which is driven by the features created by one-hot-encoder as it will only process numerical representation of any domain. As an input to the k-means algorithm, we used first sentiment analysis, emotion analysis, category and sub-category which were first encoded by the one-hot-encoder then processed to the k-means algorithms. The main objective of the utilization of k-means is to create a metadata representation on the macro level for the news dataset so we capture the overall characteristics of news articles in terms of categories and semantic analysis. We used a dynamic method to set the best number of clusters utilizing (silhouette score) then fed the k-means algorithm by the feature output of one-hot-encoder to assign clusters and create metadata representations.

For the sentiment and emotion analysis, the target was to create a global representation for each news article by analyzing the sentiment whether positive, neutral, or negative, and to have an understanding of the emotion behind the text, such as happiness, sadness, fear, anger, etc., as essential information to be captured and understood for each news article as it will enhance the distinct feature representation. As mentioned above, these features were encoded and passed to K-means to create a meaningful clustering.

Finally, the most important feature creation technique used here is the pre-trained BERT model, as it is intended to handle text data while being aware of the context, which is crucial for our use case. We have used BERT as is but did not give the needed level of distinctive ability to create meaningful embeddings; hence, we fine-tuned it to handle news articles in a much better way, which was proven by creating more distinct embeddings for news representation. There was a potential improvement here by re-training the tokenizer itself, but after deep exploration, we noticed that the tokenizer was already trained on general English language such as news articles, books and general information on websites. Therefore, retraining the tokenizer would not yield better results but more complexity to the model as we have to reset the parameters of pre-trained BERT such as maximum words domain; the default is 32,255, but if the tokenizer changed, that would be different, and might require further optimization for the BERT parameters or re-training. We proceeded with pre-trained tokenization and fine-tuning of BERT, which was enough to create distinct embeddings to represent news articles while being aware of the context.

We used sentiment and emotion pre-trained models to analyze news articles and create features of semantic meaning, alongside with the existing data of category and subcategory all of these four metadata information were passed to on-hot-encoder to be encoded thus being ready for k-means algorithms to create meaningful clusters. This was the first created embedding at size of 304 and the second resulted from the news articles processing through fine-tuned BERT at size of 768. Both were concatenated and padded to create single embedding representations for each news article. This approach allowed us to create meaningful embeddings that could be analyzed and learned through the downstream steps of the code.

Utilizing these embeddings, we have created a user profile for each user containing both preferences and none-preferences, for each user. Those profiles were created based on the average of all news articles that a user has interacted with such that for those news clicked by the user all relevant embeddings were averaged to have the best representation of user preferences profile in a single embedding, same with the none-preference profile. Then , those profiles were checked by cosine similarity to verify that they are distinctive enough to make predictions. So, during inference making, every single news article embedding will be compared to the ground truth of user preferences and none-preferences profiles as embeddings to embedding similarity evaluation, and the highest alignment score will be predicted in alignment with its relevant meaning. i.e., the similarity score for any news article with a preferences profile is higher than the non-preference profile. In that case, the user will be more likely to click the prediction. This approach scored around 87% accuracy during validation while making the model simple and very explainable as well as interpretable. The core idea of this code is to focus on creating well-structured and accurate representations of news articles by applying the best practices for such use cases of news recommendation systems while keeping in mind the end target of the best-personalized content experience.

Development

After multiple runs conducted, more than 1,500 hours of local and cloud-based run time, more than 500 experiments, the team's efforts have come to fruition to deliver a near-perfect personalized news recommendations for all users in our database with outstanding accuracy to reflect their liked, disliked, and neutral news patterns. The team's development process started by employing a multi-stage process to create a comprehensive news recommendation system. The process begins by fine-tuning BERT model for masked language modeling on news content, which includes 'title,' 'abstracts,' 'category,' and 'subcategory.' Then, the model generates dual embeddings: BERT embeddings capture the semantic content of news articles, while K-means clustering provides additional categorical features. These features are then concatenated and padded to create a unified representation for each news article. Our system then constructs user profiles by analyzing user interaction data, creating both preference and non-preference profiles based on clicked and non-clicked news articles. These profiles are generated by averaging the embeddings of articles users have interacted with, creating two distinct vectors that represent what users typically engage with and what they tend to avoid. The approach also includes sentiment and emotion analysis using specialized models to enrich the feature set. The effectiveness of these user profiles is evaluated through cosine similarity metrics, which measure how well the profiles distinguish between a user's preferred and non-preferred content, enabling personalized news recommendations based on these learned preferences.

After finalizing our trained model, the validation approach involves several key steps to evaluate the effectiveness of the trained recommendation system. First, it processes a behavior dataset containing user interactions with news articles, standardizing the format of user IDs and news IDs while ensuring proper separation of concatenated IDs with commas. The validation then uses the trained models and user profiles to assess prediction accuracy. It evaluates the system's ability to predict which news items users won't click on by comparing the model's predictions with user behavior. The validation employs cosine similarity metrics to measure the alignment between user preference profiles and news article embeddings, processing the data in batches for efficiency. The system also incorporates parallel metric calculation for faster evaluation. We used DagsHub to manage MLflow and Weights & Biases (WandB) tracking for various performance metrics, including precision, recall, F1 score, and accuracy. This validation process helps assess how well the model captures user preferences and predicts user behavior, with results being logged and monitored through modern tracking tools for thorough performance analysis.

The testing approach builds upon the validation process to assess the model's performance on a separate test dataset. The testing code implements a more rigorous evaluation of the news recommendation system by processing behavior data in batches and using parallel processing for efficiency. It focuses on predicting whether users will click or not click on displayed news items based on their preference profiles. The system uses cosine similarity metrics to compare news embeddings with user preference and non-preference profiles, making binary predictions for each news item. The testing process includes comprehensive tracking of successful and failed predictions, maintaining counts of found and not-found news IDs, and storing all predictions in a structured format. The results are saved to a CSV file with the original display list and predictions for both clicked and not-clicked news items. This testing phase provides a real-world performance assessment of the model's ability to predict user behavior and helps identify areas where the system might need improvement. The process mirrors actual deployment conditions while maintaining efficient batch processing and parallel execution for large-scale test datasets.

After receiving excellent prediction results for each user in our dataset, we deployed our personalized content recommendation system using Streamlit and deployed it on Google Cloud. The application is accessible using the adjacent QR code. The interface allows users to put in their user ID, and if they are existing users, the system will welcome them back and start serving them personalized news. Human interaction can be dynamic, and exploring new content based on certain circumstances is human nature. For example, suppose user U802334 prefers sports, auto, and lifestyle, and this day is election day. In that case, the user can navigate to another tab to view the politics category and start reading as he wishes. Additionally, five buttons below each generated news article appear for the user to select from: read, like, dislike, neutral, and skip. Once the user chooses any of these buttons, each interaction will be logged to our interface and exported as a file. The interactions will enable us to create a feedback loop for all users to better suit their preferences with their favorite news and aid us in not showing news that they don't like.

Our data includes an enormous number of records for each user. Therefore, to ensure continuous integration and deployment, scalability, collaboration, and version control, we utilized Google products to support the back-end of our product, as it will enable us to leverage its services to our needs. Our training, validation, testing, deployed model, and versions are containerized in Docker and deployed in the Google

Cloud Platform (GCP). We also utilized Google Cloud Storage and BigQuery for storing our large datasets and integrated BigQuery with tableau to allow for effortless data visualization and reporting.

The diagram below showcases our methodology and approach to tackle the challenging MIND dataset:

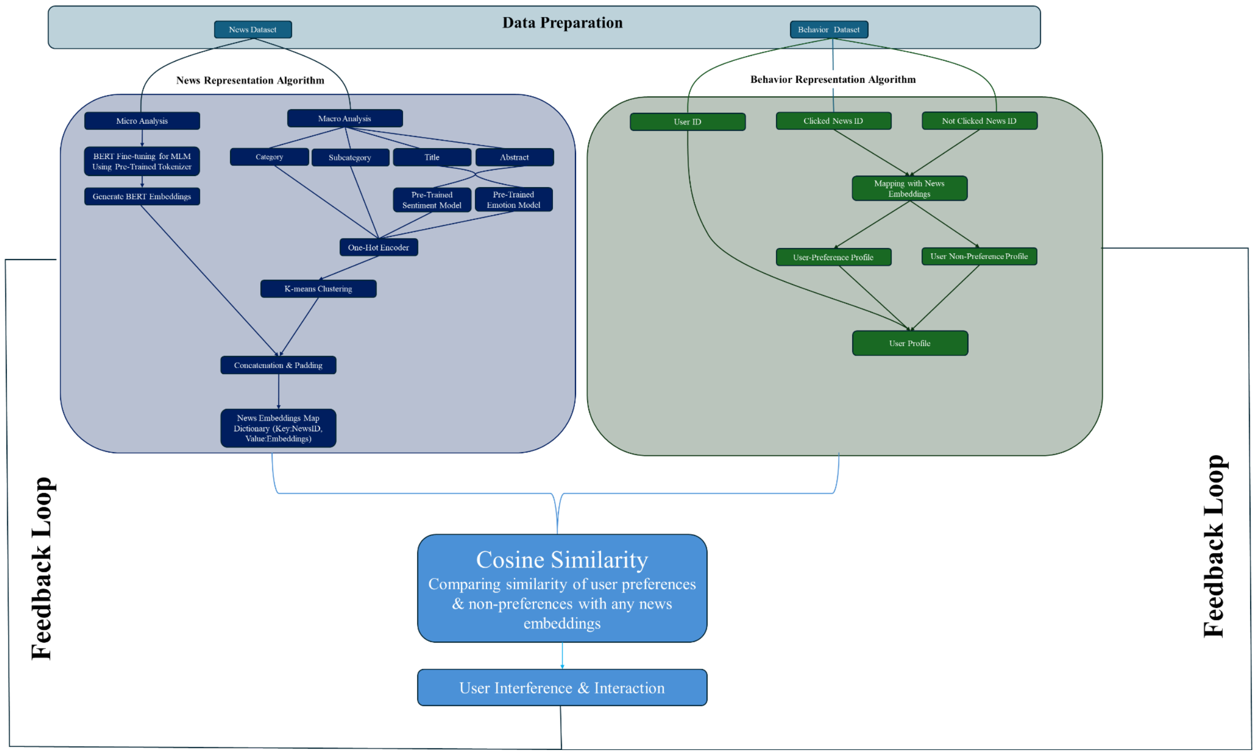


Figure 15: Data Preparation, Workflow and User Interface

Tool integration

Working on a project of this scale can be complex, particularly when working on a group project remotely. We reviewed various tools to manage this project, integrating key components such as our Python code, GitHub, and MLFlow integration.

Feature	**DagsHub**	**WandB**	**Neptune.ai**	**MLflow**	**TensorBoard**
Git Integration	Excellent	Good	Limited	Requires Setup	N/A
MLflow Integration	Built-in	Integration available	Integration available	Itself	Limited
Data Versioning	DVC Integration	Limited	Limited	Requires DVC	N/A
Visualization	Good	Excellent	Good	Basic	Excellent (for TensorFlow)
Collaboration	Good	Excellent	Good	Limited	Limited

Cost	Free tier	Can be expensive	Free tier	Open-source	Open-source

Having conducted the analysis summarized in the above table, we opted for DagsHub. DagsHub was simple to integrate with our code and streamlined the training, validation, and testing phases. Therefore, the team has met with the CEO of Dagshub to get his insights on integrating our code, data, and experimentation. Throughout this journey, Dagshub has proven its capability to seamlessly integrate all of our project data in a scalable and simple way.

Data Architecture

Figure 16 below illustrates our project's data architecture. This is broken down into three major components: Model Development, Application Components, and Application Deployment.

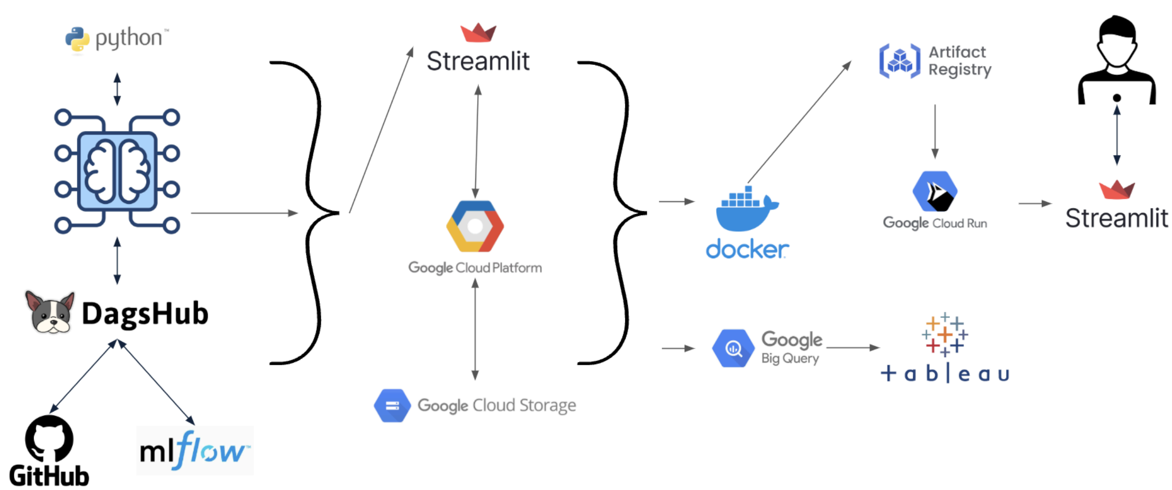


Figure 16- Data Architecture Diagram

Financial Analysis

The deployment of the personalized recommendation system demonstrates a robust financial business case, signifying considerable potential for profitability and the creation of enduring value.

A five-year projection indicates that the system will necessitate an initial investment of $800,000 for development, integration, and training expenses. Annual operational expenses are anticipated to total $150,000, including cloud hosting and continuous model training.

With an 8% discount rate and an annual net cash flow of $450,000, the project generates a Net Present Value (NPV) of around $985,000 over five years, indicating a significant return on investment. The anticipated payback period for the system is just under two years, with profitability achieved swiftly.

Aramco's digital publications do not generate direct revenue from advertising or subscriptions; nevertheless, maintaining customer engagement on corporate websites and publications is advantageous for its long-term reputation and brand value, estimated by Kantar BrandZ to be $107.7 billion in 2024[2]. The recommendation system enhances user engagement and facilitates their understanding of Aramco's values, initiatives, and expertise. This familiarity cultivates trust and preference, which are crucial for establishing joint ventures (JVs), promoting cooperation, and enhancing Aramco's global standing.

The stock market success of Aramco and investor confidence are contingent upon its reputation for reliability and stability. Aramco's stock price has demonstrated the market's recognition of its leadership and robust reputation since its initial public offering. Digital engagement enhances Aramco's reputation as an innovative, forward-looking company that prioritizes transparency and stakeholder communication.

Investors and partners generally assess a company's innovation and engagement through its digital footprint. The system technique allows Aramco's platforms to exhibit their market leadership. By increasing familiarity and preference, the system indirectly raises the company's market value, yielding financial and reputational benefits that surpass user engagement measurements.

This analysis highlights that the system functions as a strategic asset for improving engagement and brand recognition while also serving as a financially sound investment for Aramco, yielding concrete economic benefits and operational and marketing advantages.

Sustainability

Sustainability is an integral part of our personalized recommendation system. As of November 2024, the average website produces as much as 1.76g of CO_2 per page view[3]. The average user spends around 10 minutes daily reading 10-20 articles to find relevant content[4]. This produces approximately 17-35 g of CO_2 per user daily from inefficient news browsing. Users can find relevant news faster and more efficiently using our personalized news recommendation system. Users can find their tailored content within 5-7 articles, which could reduce CO_2 emissions by approximately 60% (7-14)g of CO_2.

For our platform, which includes 1 million daily active users, this can entail saving 10 - 21 metric tonnes of CO_2, or 3,650 - 7,665 metric tonnes of CO_2 per year. To put these numbers into perspective, one tree absorbs around 0.021 metric tonnes of CO_2 (10 kg) annually[5]. Therefore, through our personalized recommendation system, we could plant 173,810 - 365,000 trees annually. Additionally, the average vehicle emits about 4.6 metric tons of CO_2 per year[6], which, based on what we are saving in CO_2 emissions, will be able to take 793 - 1,666 cars off the road for a year. Lastly, the average household emits about 7.5 metric tons of CO_2 annually[7]. Therefore, we will be saving about 487 - 1,022 emissions from homes. These equivalences help demonstrate how the small efficiency gains in digital content delivery can accumulate to create significant environmental impact when scaled across millions of users.

Risk Analysis

Some risks are associated with any project and can range in severity. We, therefore, developed a risk analysis for our project to help us identify and mitigate the risks. The table below summarizes the potential risks and our strategy to mitigate them:

Type of Risk	**Risk Description**	**Severity**	**Impact Description**	**Proposed Mitigation Strategy**	**KPI for Monitoring**	**Tolerance Range**
Technological Risk, Operational Risk	System failure of performance degradation during peak usage times, leading to system downtime	High	1- User Experience Degradation: Poor recommendations leading to decreased user engagement. 2- Service Distribution: System outages affecting news delivery. 3- Resource Waste: Inefficient computing resource usage during high traffic.	1- Implement robust load balancing and auto-scaling. 2- Regular system performance monitoring. 3- Maintain backup servers and redundant systems	1- System uptime percentage. 2- Response time for recommendations. 3- Error rate in recommendations/	1- 99% system uptime. 2- Response time under 20s. 3- Error rate below 1%.
Compliance Risk - Data Privacy Risk	Potential breach of user data privacy or failure to comply with data protection regulations (GDPR, CCPA).	High	1- Legal Consequences: Regulatory fines and penalties. 2- Reputational Damage: Loss of user interest and trust. 3- Financial Impact: Cost associated with legal proceedings.	1- Regular privacy audits, 2- Implement robust data encryption 3- Regular staff training on data protection	1- Number of data privacy incidents. 2- Compliance audit scores. 3- Data encryption coverage.	1- Zero data breaches. 2- 100% compliance with regulations. 3- 100% encryption of sensitive data.

Strategic Risk - Market Risk	Risk of the recommendation system not meeting evolving user preferences or market demands.	Medium	1- User Retention: Decreased user engagement. 2- Competitive Position: Loss of market share to competitors. 3- Revenue Impact: Reduced advertising effectiveness.	1- Regular user feedback collection. 2- Continuous model retraining with new data. 3- Market trend.	1- User engagement metrics. 2- Click-through rates (CTR). 3- User satisfaction scores.	1- Minimum 80% user satisfaction 2- CTR above industry range. 3- Monthly engagement growth of 1%.

Data Governance

Data governance is a crucial component of the infrastructure for personalized news recommendation systems. To address privacy concerns and comply with legal requirements, the team has developed a detailed scheme for the MIND dataset, which contains anonymized information on over a million individuals. This academic approach to data management ensures the protection of user privacy while enabling the effective operation of these personalized news services. Our approach is similar, where we will ensure transparent data collection by clearly stating our purpose is that these data are collected to improve our recommendation system and not to be shared to other entities. The proposed framework offers clients the choice to assent or decay information collection based on their inclinations. For clients who assent, the information will be safely put away in a cloud framework with encryption conventions for both in-transit and at-rest information. Moreover, a devoted information security officer will be alloted to supervise all methods, keep up occurrence reaction conventions, and conduct standard information assurance affect appraisals. This vigorous information administration system points to guarantee client certainty, information security, and adherence to worldwide legitimate prerequisites. Through comprehensive record-keeping, compliance reports, and intermittent reviews, the framework sets up a solid and secure environment for news suggestion systems.

Conclusion

Metrics

After multiple trials through the training code versions we built, the team has successfully reached an optimal level of excellent results. We used the confusion matrix to evaluate our prediction in all training versions, and we got the best results, as shown below, which are attributed to the implemented model:

Overall Evaluation Results

Precision	**Recall**	**F1 Score**	**Accuracy**
0.8760	0.9421	0.9038	0.8760

These metrics resemble the team's hard work, dedication, and aspiration to achieve an excellent prediction to generate content for users based on their preferences.

Impact and Business Implementation for Boosting Brand Recognition with AI-Driven Content Suggestions

In today's digital landscape, owned channels, such as corporate websites and digital magazines, are crucial for establishing a company's brand identity and engaging its audience. For Aramco, these channels directly communicate with stakeholders, providing control over messaging and facilitating personalized content distribution. By leveraging content on owned channels, Aramco can improve user engagement, acquire significant data-driven insights, and reinforce its market position, securing a sustainable competitive advantage.

Business Impact

The customized news recommendation model developed for Aramco can revolutionize the company's engagement with its digital audience. The system employs sophisticated machine learning techniques to provide personalized content that enhances user experience, maximizes marketing efficiency, and yields substantial financial outcomes. By allowing the model to adapt the displayed content across its publications depending on the user's preferences, we are aiding consumers' understanding of the company's activities and initiatives. This familiarity fosters confidence in Aramco and enhances Aramco's reputation with key stakeholders, ensuring its position as a preferred partner in global partnerships and joint ventures.

In addition to improving the user experience, we are creating additional value for Aramco. The model also allows us to extract audience clusters based on their content consumption on the Aramco website. This, combined with first-party cookies, will allow us to target these audiences with tailored advertising even when they are not on Aramco-owned channels.

The technology analyzes user behavior and preferences and produces actionable insights that help Aramco refine its content strategy. Data-driven decisions enhance the pertinence and efficacy of Aramco's marketing strategies.

Feasibility for Business Implementation

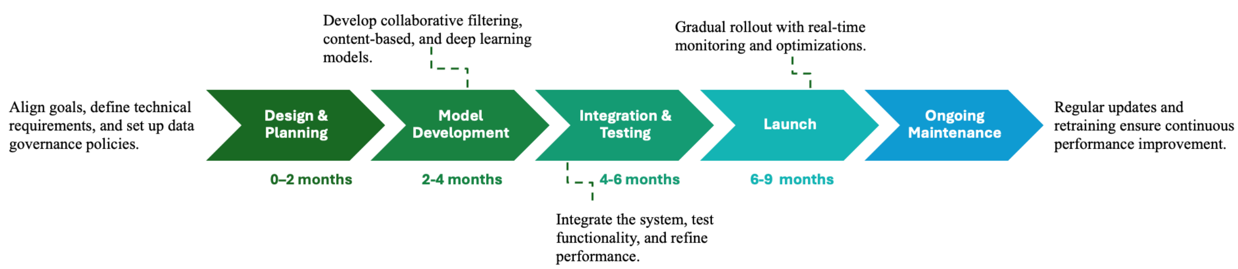
Implementing this recommendation system is feasible and achievable for Aramco, owing to its scalable architecture, capacity for smooth integration, and emphasis on ethical and legal compliance. This is why it is both logical and significant for the organization.

The solution is engineered for effortless integration into Aramco's current digital framework. The recommendation engine can be effortlessly incorporated into the company website, AramcoLife, and Elements Magazine without necessitating substantial alterations. APIs enable seamless interaction between the recommendation engine and current platforms, guaranteeing clients receive customized information swiftly post-implementation. This reduces disturbance and enhances efficiency throughout the integration process.

The system's primary advantage is its capacity to scale with Aramco's expanding digital presence. The cloud-based design enables rapid adaptation to a growing user base or the integration of more publications and content categories. The technology accommodates an expanding user base and can be integrated with additional data sources without modifications to the design. The system is precisely engineered to prioritize data security and adhere to regulatory compliance. It complies with the highest standards to protect consumers' data, including GDPR. The recommendation algorithms have been assessed for fairness to reduce biases in content distribution, guaranteeing equitable treatment for all users. This emphasis on ethics and compliance mitigates legal risks and improves user confidence.

Implementation Roadmap

To ensure the effective execution of this personalized recommendation system, we have established a comprehensive roadmap covering the stages from system design to complete deployment.



By following the proposed implementation roadmap at Aramco, we can deploy and refine this system to meet the evolving expectations of its audience. Through strategic execution and ongoing optimization, this solution will support Aramco’s digital transformation and reinforce its leadership position within the global energy sector. This project underscores the importance of leveraging advanced technologies to drive innovation and sustain competitive advantage in a dynamic industry landscape.

References

1. Isaksson, R., & Suljanovic, M. (2006). The IKEA experience : a case study on how different factors in the retail environment affect customer experience (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-59251

2. BrandZ Most Valuable Global Brands 2024 - https://indd.adobe.com/view/c21c3b44-92cb-4386-848e-c1641347979b

3. Root Webdesign - How much CO2 does a website produce? https://rootwebdesign.studio/articles/how-much-carbon-does-a-website-produce/#:~:text=The%20average%20website%20produces%20as,of%20CO2%20per%20page%20view.

4. Taglab - Average Time Spent per User Metric Definition https://taglab.net/marketing-metrics/average-time-spent-per-user/

5. How much CO2 does a tree absorb? https://www.viessmann.co.uk/en/heating-advice/boilers/how-much-co2-does-tree-absorb.html

6. EPA - Greenhouse Gas Emissions from a Typical Passenger Vehicle https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle#:~:text=typical%20passenger%20vehicle%3F-,A%20typical%20passenger%20vehicle%20emits%20about%204.6%20metric%20tons%20of,8%2C887%20grams%20of%20CO2.

7. CCFPD - Facts Chart https://www.ccfpd.org/portals/0/assets/pdf/facts_chart.pdf

Appendix

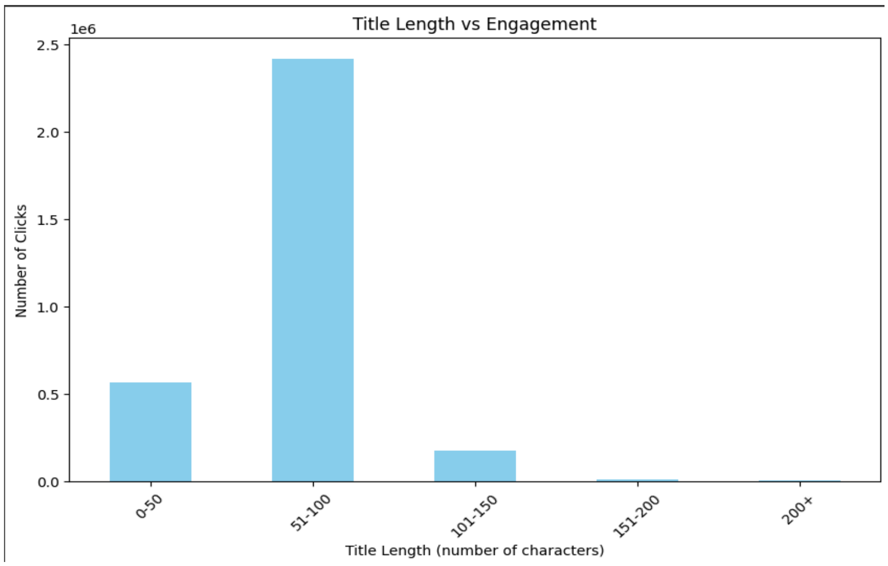


Figure-3: Title Length vs Engagement

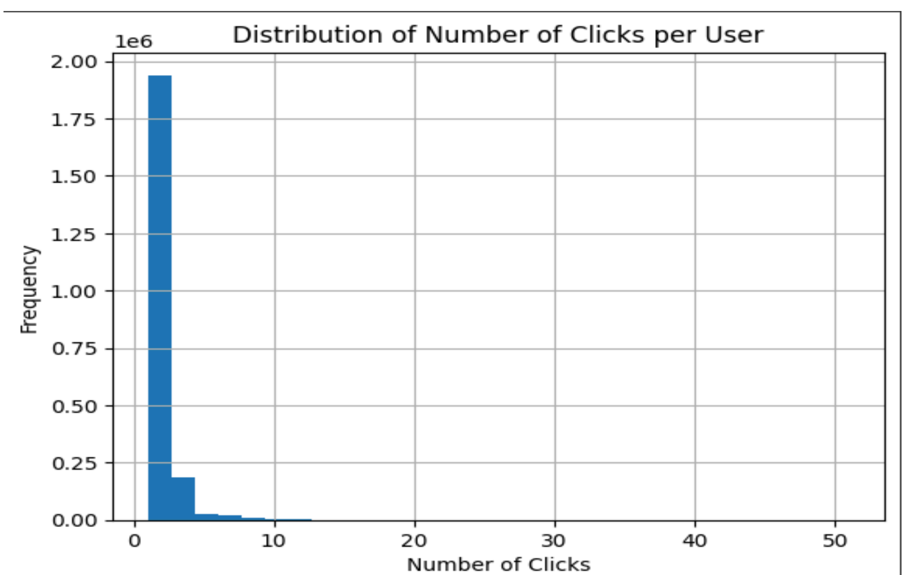


Figure-4: Distribution of Clicks per User

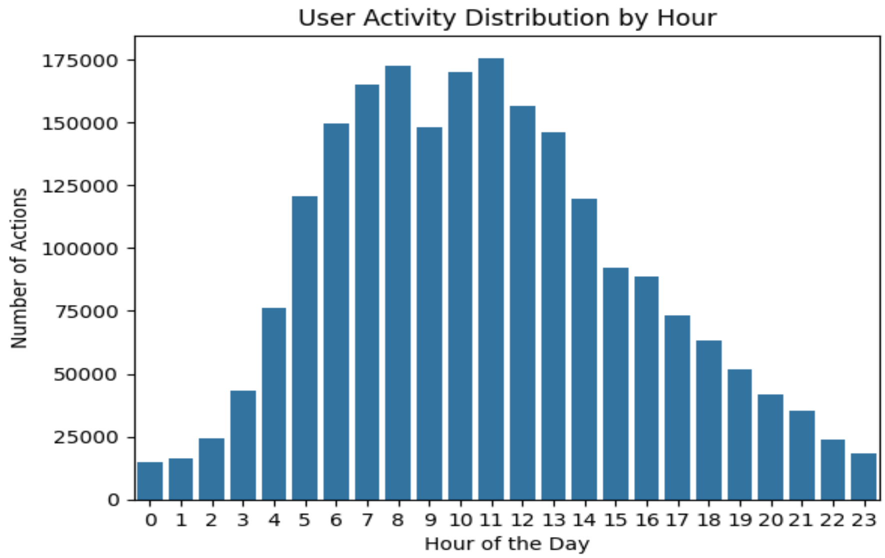


Figure-5: User Activity Distribution by Hour

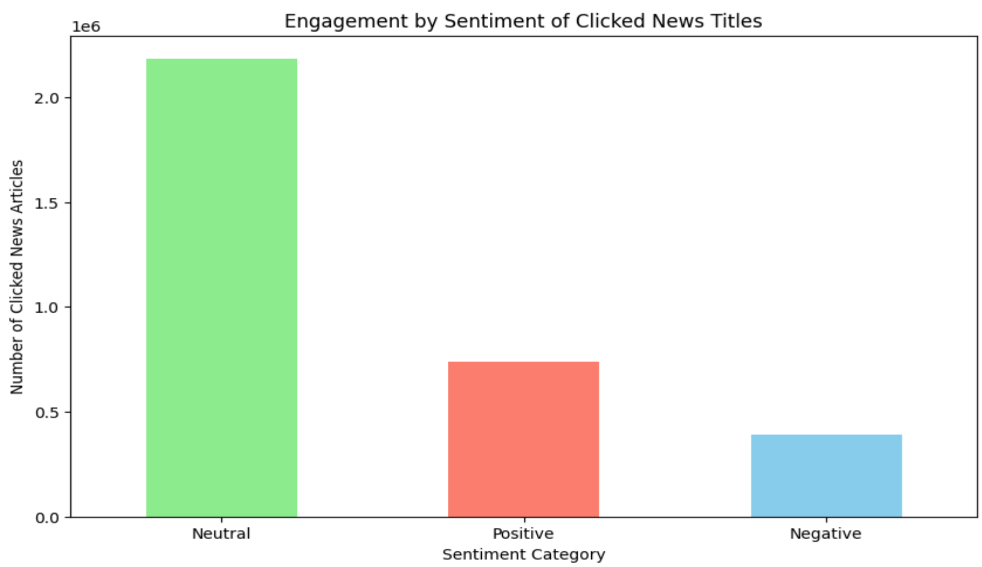


Figure-12: Engagement by Sentiment

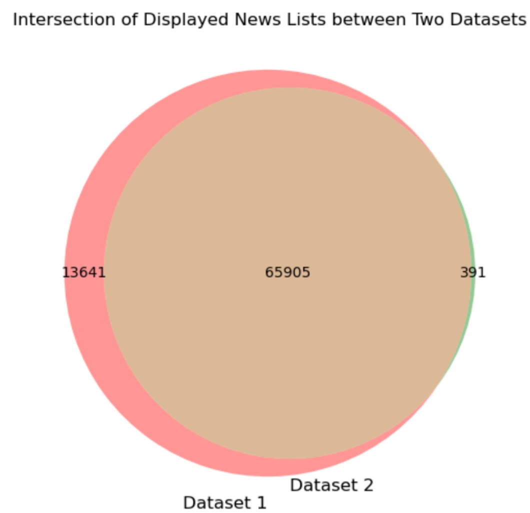


Figure-13: Training and Validation News Overlap

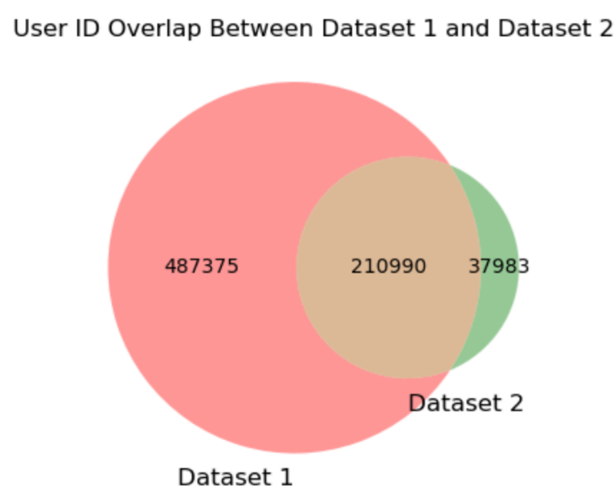


Figure-14: Training and Validation User ID Overlap