TIEVS – CAR CLASSIFIEDS RECOMMENDATION SYSTEM

2021-195

Project Proposal Report

R.A.D Prathapa – IT18122060

Supervisor – Ms. Manori Gamage
Co-Supervisor – Ms. Suriyaa Kumari

B.Sc. (Hons) Degree in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2021

TIEVS – CAR CLASSIFIEDS RECOMMENDATION SYSTEM

2021-195

Project Proposal Report

R.A.D Prathapa – IT18122060

Supervisor – Ms. Manori Gamage
Co-Supervisor – Ms. Suriyaa Kumari

B.Sc. (Hons) Degree in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

March 2021

DECLARATION

I declare that this is my work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning. To the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student ID	Signature
R.A.D Prathapa	IT18122060	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date:

ABSTRACT

This proposal in aimed at developing a customer-optimized recommendation system within the main classified advertising platform, 'Tievs', to provide the customers more conveniency and ease of use, when extravagating through the application for the necessary preferences in a timely manner. The common problems in classified advertising platforms are that when customers find it difficult to finding their desired items and exit the application all together, without wasting time, thinking that the item is not currently present, although the item is genuinely present, deep inside the search lists, without being prompted. Some recommendation systems have already been implemented in the classified platforms already in market, but the issue remains same, which demonstrates the internal recommending process has not been enhanced meticulously. Moreover, previously conducted research have not made their target as assessing the recommendation functionality within classified advertising applications, using machine learning appliances. Sparsity problem, Cold-Start problems are common in collaborative filtering techniques, which requires user inputs such as likes, click counts, comments, feedbacks etc. to proceed further with recommendations, although those are rarely found in classified ad applications. Thus, it is an objective of this proposal to investigate content-based filtering techniques to clarify the most relating advertisements for the customers, since the above stated user dependent attributes can be ignored in the implementation processes. In addition, ensemble models have been rarely exercised to imply with user suggestions, with regard to content-filtering as well, which is another aspect neglected in its subjective research field, that this proposed system will be addressing. In conclusion, a more personalized, revenue increasing, user-satisfaction empowered, efficient advertisement discovering recommendation engine will be built using TF-IDF and cosine similarity measures to support the proposition of amplifying the combined model's accuracy in recommending classifieds to customers.

Key Words – recommendation, TF-IDF, cosine similarity, ensemble.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1 INTRODUCTION

1.1 Background & Literature Survey

Classified ads are a market that allows users to sell their used/unused items or a service by posting advertisements. The market is quickly gaining traction for online classified ads. People have been less eager to utilize traditional classified advertising methodologies like newspapers; hence the recent trend and popularity of online platforms providing the customers the necessary satisfaction, only more improved than the old ways of posting or viewing classifieds of various types of products. Newspaper revenues from advertising classifications are continuously declining as Internet classifications are growing.

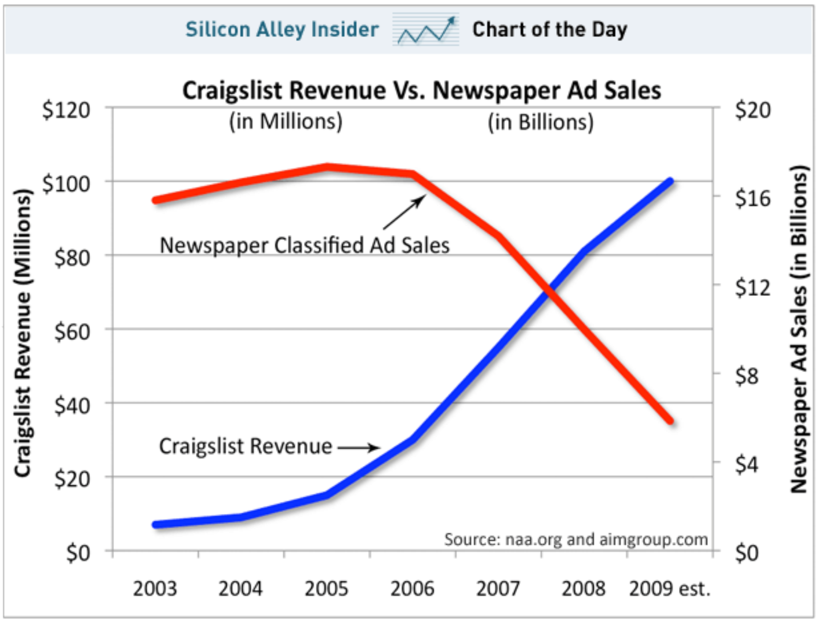

Figure 1.1 – Craigslist Revenue Vs. Newspaper Ad Sales

The benefits of having an online platform for classifieds can see as below, and the customer is the one who seeks out the classifieds when they need a particular product. Potential customers will be able to search for the item or service all the time. It is effortless to search the relevant well-sorted product on the web platform rather than searching in the newspaper. Another benefit is that the user can upload images videos about the item to increase the advertisement's reach [1]. Due to those benefits, in figure 1.3, from 2005 to 2009 use of online classified ads sites has been doubled. As shown in figure 1.2, in the UK, more people prefer online websites rather than going directly to a car dealer when they need to buy a used car.

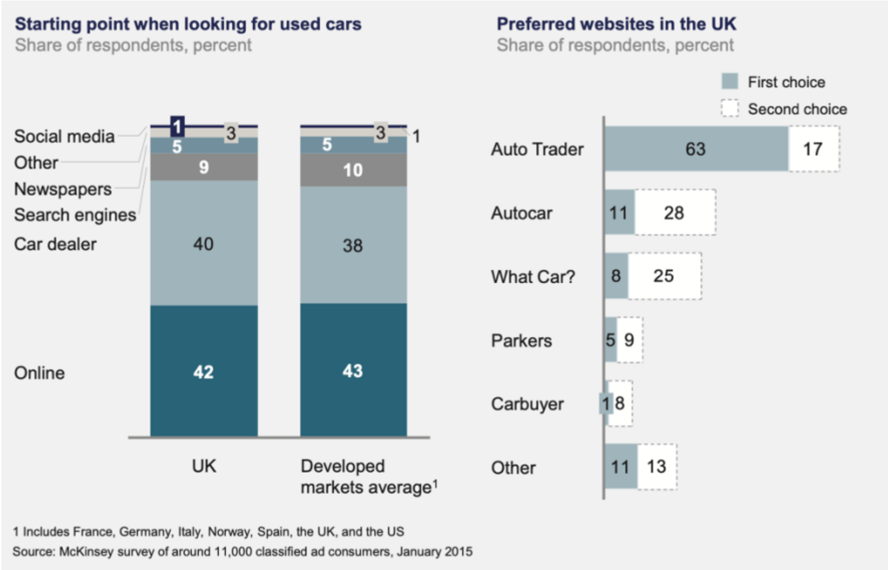

Figure 1.2 - Starting point when looking for used cars

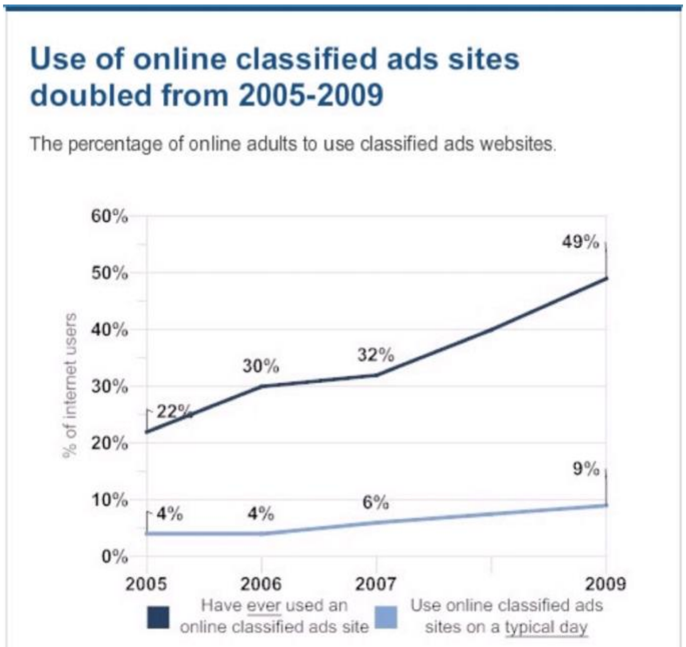

Figure 1.3 - Use of online classified ads sites doubled from 2005 -2009

With the World Wide Web's growth, the website's complexity and size are also on the rise. Therefore, it is challenging, complicated, and time-consuming to find information, data, or elements on those websites. When the domain is quite large, it is difficult for users to make an appropriate decision. In this regard, a recommender system provides the users with a refined list of alternatives tailored to their preferences. With a recommendation system, the website can recommend the items that the users have an interest in and the items they are searching for. The recommender system can predict whether a particular user would prefer an item or not So that the recommender system can deliver relevant items to the customer. Customers will be more engaged with the site when personalized product recommendations are made, thus boosting communication between them and the system. These exciting items may be any textual information, and it may also be an index on a particular topic that the user is searching for. A website can be recognized as a compilation of the elements involved. Within a recommendation system, we have an item and a user. To recommend the product to the customer, recommend system requires specific parameters like rating or item attributes. Competition is growing day by day among online retailers, and recommendation systems are a way of personalizing a person's needs. Recommendation systems are like a salesperson in a web store.

They are mainly five types of recommendation systems as figure 1.4 depicts,

Content-Based: This system recommends items that are similar to other items that the user has liked in the past. The similarity of the items is determined on the basis of the characteristics associated with the items being compared [8].

Collaborative Filtering: This system is called as "people-to-people correlation." Collaborative filtering is considered to be the most popular and widely implemented technique in the recommendation system. It works on Neighborhood methods, which are focused on relationships be-tween items or between users.

Demographic: This system recommends items based on the demographic profile of the user.

Knowledge-Based: This system recommends items based on specific domain knowledge about how certain item features meet users' needs and preferences of the user. This system will work better than others at the beginning. Nevertheless, if they are not fully equipped with learning components, then they may fail.

Hybrid: Combine one of two systems to a specific industry, for example, content-based and collaborative filtering.

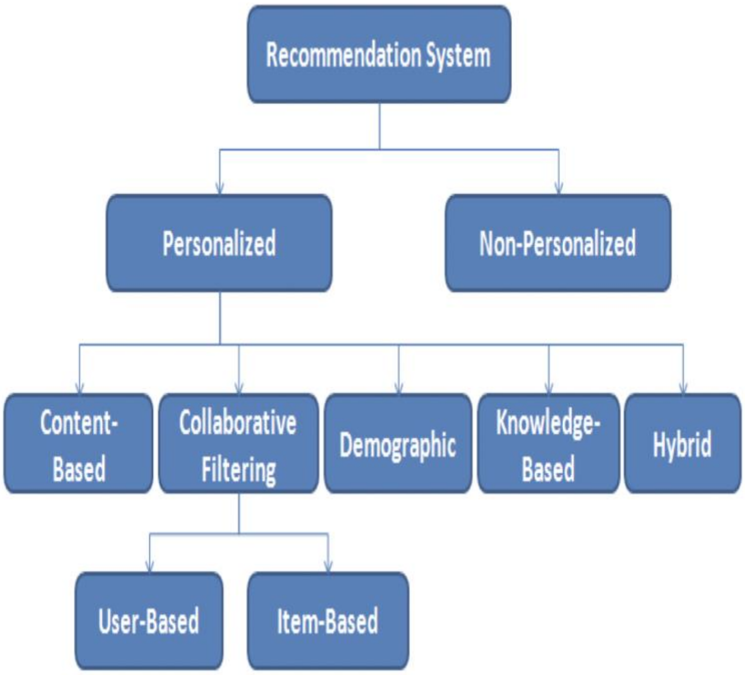

Figure 1.4 - Types of Recommendation System [7]

In recent years, many forms of study have been carried out for recommendation systems. They have used different algorithms and different types of recommendation concepts such as content-based, collaborative, hybrid methods. There are many types of research for a system like a movie, social networks, music recommenders, e-commerce related systems.

R. Singla, S. Gupta et al. have done FLEX, a content-based movie recommender system [3]. They have used Distributed Bag of Word's version of Paragraph Vector

(PV-DBOW), Term Frequency algorithms for the recommendation engine. With this, they have analyzed movie plots/ descriptions and generate an inter-movie similarity score. This method is inspired by the 'Paragraph Vectors' doc2vec model approach.

Another research has been done by A. Pal et al. regarding an improved content-based collaborative filtering algorithm for movie recommendations [4]. They used a modified hybrid approach that includes content-based and collaborative-based approaches. Their dataset contains a total number of 10004 ratings, 9125 movies, and 671 users. As in figure 1.5, tags and genre take into considerations. First, various users' tags for and movie are used in methodology and translated into a single list. The genres for each movie are appended to the same list of tags. This final list is referred to as the objects for a particular movie. The object set for each active movie is compared to the object set for each other in the dataset, and the number of matching objects is assigned to the set. Then using a formula similar movie list is formed.

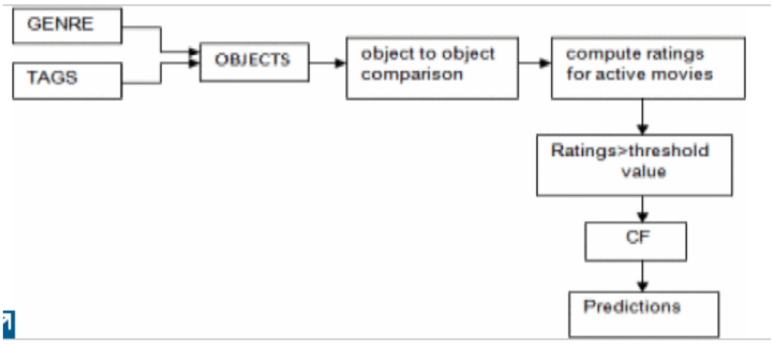

Figure 1.5 - Flowchart of hybrid collaborative filtering algorithm

According to the author, B. Kostek, the challenge in search solutions lies in their scalability. In his research paper [5], he mentioned music recommendations using collaborative filtering. When one user listens to preferred music, the system will find other users with similar music tastes then the system will recommend the first user with new songs that he/she might like. User, Item, and Rating are the basis of the collaborative filtering. In figure 1.6, green lines show the user's choice of songs, and a red arc shows whether two users are interconnected by a given song, with the violet dashed line show the recommendation.

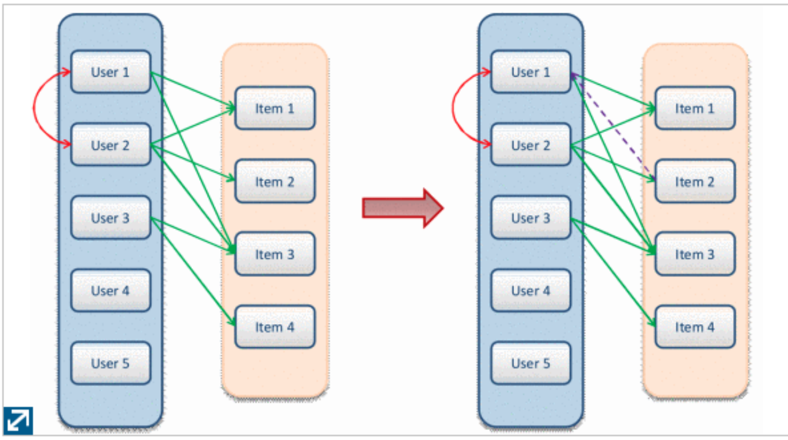

Figure 1.6 - User-based music recommendation

Author S. Shaikh has used a graph-based approach for a recommendation system in e-commerce [6]. This proposed system uses an overlap semantic method to integrate recommendations and semantics. Data will be stored in a graph format, and the Overlap value will be calculated using the overlap formula. Only items that are considered for a recommendation have a correlation value of more than 0.4. The process is shown as an outline in the below figure 1.7.

$$\text{Overlap} = \frac{N(p \cap q)}{\min(N(P), N(q))}$$

View Source

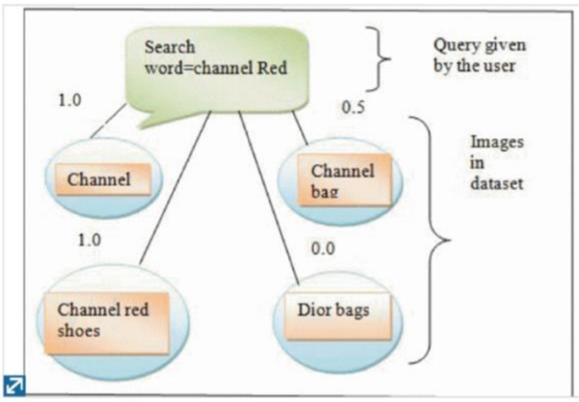

Figure 1.7 – Overlap formula and working model

1.2 Research Gap

There are many types of research for a system like a movie, social networks, music recommenders, e-commerce related systems. Still, there is not enough research for classified ad recommendations, let alone car advertisements. According to the information that we can extract from other research, the following attributes to build the recommendation has been considered.

- votes
- likes/dislikes
- ratings
- views
- feedbacks
- comments

According to the available research papers [3]-[7], they required user preference attributes as above. Those attributes may not be included in a classified advertising application almost in many scenarios hence the monitoring for such attributes are in vain. Besides, it creates the Cold Start problem, which is information unavailability for the recommendation engine to process the newly added component. The proposed system would not be accounting such attributes, except for content based features, to avoid the Cold Start problem. Most of the analyzed research [3]-[7] have not considered a sufficient number of records included dataset, having content characteristics, to train and test their relevant models for recommendation, although existing classified advertising platforms in market monthly accepts more than 200,000 advertisements. Hence the usage of a large dataset is more convenient since the learning range will then be expanded further, which the proposed recommendation system will be targeting to achieve. Not to mention, the examined research [3]-[7] have not exercised the idea of a combination of TF-IDF and cosine similarity algorithm types, to experiment the accuracy levels and overall performance. Thus, the proposed system will be using the most suitable and optimized content-based recommendation, having a combination of models, for classification of classified ad preferences for each customer. The table 1.1 shows a tabularized format of the explanation.

Table 1.1 - Comparison of former research

Research	**Optimized for classifieds ads**	**Combination of different algorithms**	**Filtering done on car characteristics**	**Addressing the cold start problem**
Research A [3]	✗	✓	✗	✗
Research B [4]	✗	✓	✗	✗
Research C [5]	✗	✗	✗	✗
Research D [6]	✗	✗	✗	✓
Research E [7]	✗	✓	✓	✗
Proposed Recommendation System	✓	✓	✓	✓

1.3 Research Problem

Several recommendation algorithms have been developed throughout the recent years, regarding movies, videos, images, economical blogs etc. However, the need of an appropriate recommendation system with a significant level of accuracy, targeted towards classified advertising haven't been conveniently fulfilled, as of yet. Most of the implemented models need user browsing history, clicks, likes for suggesting the customers with their necessary selections. These attributes caused a cold start problem in the recommendation system. The cold start problem is when a new item creates in the system, and it does not have enough data to process in the recommendation engine; therefore, the new item will not be displayed in the recommendation section. Sufficient amount of research hasn't been conducted to address the above stated problem [8]-[11].

The customers rarely give ratings, comments or feedback on classified advertisement platforms, hence considering those attributes for recommendation purposes is entirely unavailing and the probability of inaccurate predicted suggestions will increase as well. Then the problem occurs when finding what kind of content attributes should be considered in order to provide those customers with personalized preferences in a more efficient way. In addition, algorithm training, with the use of a dataset having a large number of records, such as 450,000 has never been attempted in previously analyzed research [3]-[7]. In general, approximately 500,000 new classifieds are being submitted into main classified advertising platforms such as Craigslist, Olx etc. thus, training a recommendation algorithm model based on a limited amount of record set is ineffective and problematic. To improve the accuracy further, more and more information needs to be infused into the models. Besides, it's more challenging to identify the most impactful features for car ad recommendation, and for the training algorithm models themselves as well. Moreover, the combination of TF-IDF (Term-Frequency - Inverse Document Frequency) and Cosine similarity was not used for the procedure of product recommendation [13], in the former mentioned research [3]-[7], giving rise to an uncertain hypothesis, that a better performing, scalable and a robust algorithm model is yet to build. Addressing the problematic hypothesis, is one rumination of this proposed recommendation system [8]-[11].

2 OBJECTIVES

2.1 Main Objective

The main objective of this particular proposed system, having an internally driven, optimized item recommendation process, to visualize the best and most convenient preferred advertisement results for customers, who are utilizing the 'Tievs' application, with a high precision and speed, so that the customer satisfaction is uplifted to a commendable level, giving rise to application remarkability in utilizing new technologies for the purpose of improving usability and user friendliness.

Without a thoroughly advanced and meticulous recommendation system, the necessary items that the customers might be wanting to explore, would not be presented to them, making the customer search those specific needs individually, wasting time from their own busy schedules. As a result, the application as a whole will face a high customer turnover rate, giving rise to the probability of application reputation degradation. Nevertheless, much research has been done to address the above issue, for applications containing movies, videos, images, social media contents etc. although only a few of them have targeted classified advertisement platforms. Many have used either content-based filtering types as singular algorithms, or collaborative filtering through various versions of algorithms. Collaborative filtering requires customer input attributes such as likes, clicks, comments, feedback etc. and in classified advertisement applications, the recommendation engine cannot entirely depend on them, due to their scarcity, making the prediction process of preferred items for the customers, less accurate. Therefore content-based filtering seems much convenient, regardless not being investigated further in detail as to consider the novel implementations of it, targeting the improvement of the recommendation processes, within classified advertisement platforms. Thus, the ulterior aim of this proposal to provide the customers an error-free, scalable, reliable classified ad recommendation system, all the while utilizing the content-based approach in a novel procedure.

2.2 Specific Objectives

In order to reach the main objectives, the specific objectives that needs to be attained is as follows,

1. Implementation of a combination of models for optimized classified recommendation.

Since in most research conducted, the application of a singular algorithm model, has been entertained. This proposal fundamentally exercises the content-based filtering approach, which can be known as the most appropriate type compared to collaborative filtering, that of which requires attributes such as users likes, comments, feedbacks, clicks etc. to function, although they are not commonly found in classified advertising systems often. Within the already implemented content-based recommendation systems, the usage of an individual algorithm model type, is to be seen more than often. Therefore, the attempts for ensemble models can be employed as sparse, hence a specific objective of the proposed system, to combine cosine similarity and TF-IDF.

2. Relevant Feature investigation

The dataset containing about 26 characteristics relating to cars, which were sold through the online classified platform Craigslist. Out of those 26 characteristics, the most suitable ones must be clarified, that specifically contributes to the user recommendation procedure to function with an impressive performance. Although it may seem trivial at first glance, this process can be known as the most challenging and impactful, for the final system output result. On condition of successful feature analyzation, extraction, transfer to the training of the selected algorithms, the rest of the steps to be conducted, would be less troublesome.

3 METHODOLOGY

3.1 Research Methodology

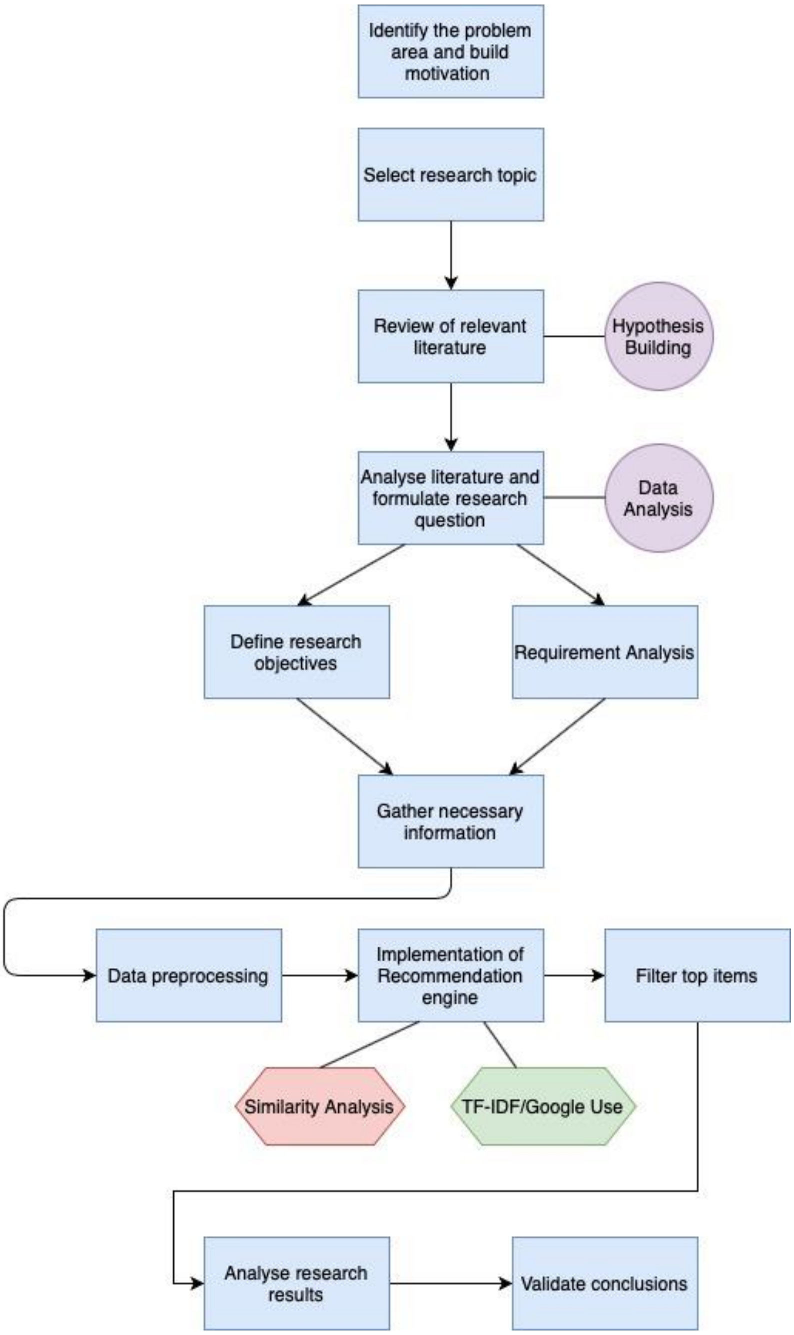

Figure 3.1 – Research Methodology

3.2 System Overview

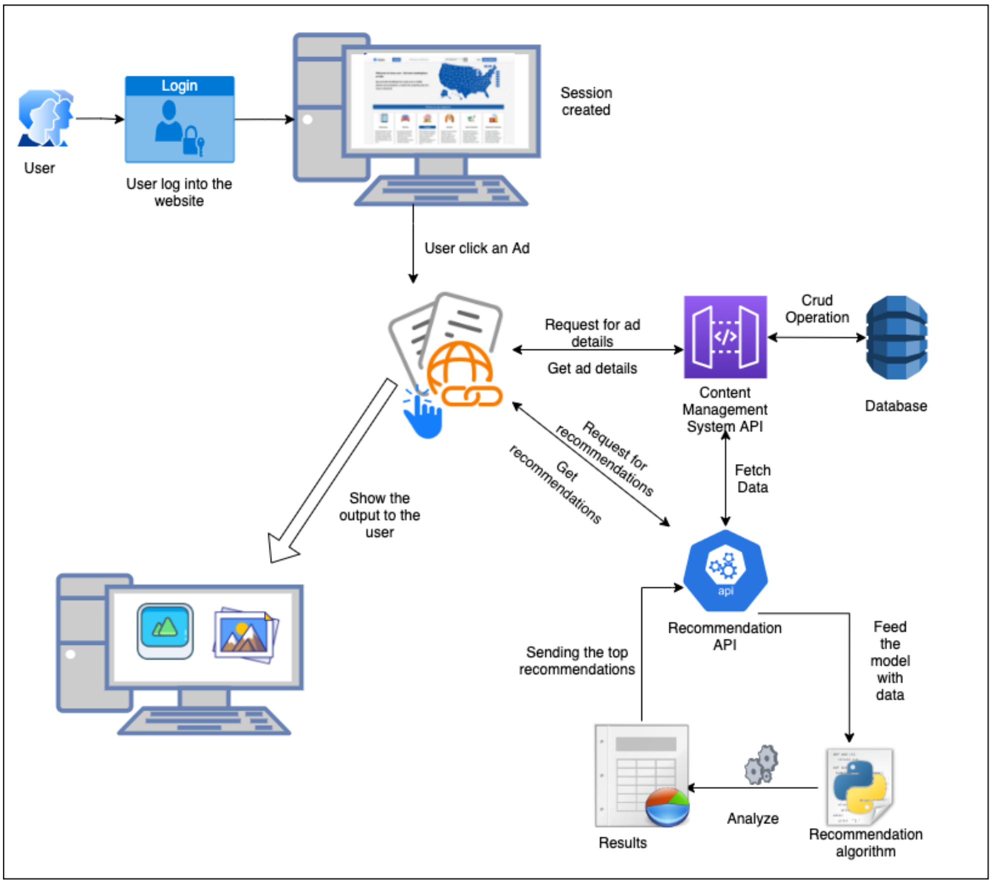

Figure 3.2 - System Overview Diagram

The above-shown Figure 3.1 is the overall high-level diagram of the proposed system. The first user logs into the system, it will create a user session, then the user clicks an ad(hyperlink), it will request the ad details from the content management system, and also it will asynchronously request recommendations from the recommendation API. The recommendation API will request other ads in that category from the content management system, and current Ad details will be analyzed using the optimized algorithm for the classifieds ad. The top recommendations will be given to the user as a response to the API request.

3.3 Software Development Life Cycle

The life-cycle development of the software that would be considered will be agile methodology. In addition, under the vast Agile framework, Scrum methodology will be mainly covered. Scrum is a lightweight framework that allows individuals, teams, and institutions to gain value through agile strategies for complex challenges. Product backlog creation, sprint planning and creation of sprint backlog, daily scrum meeting, sprint retrospectives, sprint review and next sprint planning are the incremental but iterative basic steps, that are being followed within a scrum approach followed. For this proposed system, the same procedures will be effectively followed. It is safe to state that when a research is being conducted, many unanticipated events can occur, especially since some of the would be studied areas should be analyzed for the foremost time. In that scenario, the already accepted and established plan, will not be followed in the exact same way, since when addressing the impediments of the research process, alterations will have to be made. In that way, Agile methodology is significant for conveniency of the continuous and persistent flow of the implementation process. According to the following Figure 3.2, Agile will be giving the research more control, better productivity, better quality, higher customer satisfaction and higher return on investment as well.

Figure 3.3 – Benefits of Agile Methodology

3.4 Data Preprocessing and Analyzation

The necessary dataset for implementing the proposed recommendation system will be acquired from the online dataset available web portal Kaggle. The specific dataset that is being used for this research is the ‘Used Cars Dataset’ uploaded as ‘Vehicles.csv’ by Austin Reese, which has been continuously updated monthly, staring from 2020 November. The data included in the dataset was scraped through the infamous online classified advertising web portal known as Craigslist, by using a web scraper. The dataset contains about 450,000 records of sold cars with their relevant characteristics, on the above-mentioned online platform. One of the main reasons for selecting this particular dataset is the large number of records itself, since in order to achieve a more premium recommendation of car advertisements for the customers, the more amount of information to train the necessary algorithms, is the better. As examined in the previous research [3]-[11], none of them have used a similar complex dataset for their research purposes. Besides, the data contained was updated about three months ago, and therefore much more convenient to acquire the latest details about the reselling cars. Moreover, the need for manual data collection degrades since it is more time consuming, considering the proposed scope of this research, and also a large range of information cannot be attained through that as well. Next, the data preprocessing methods must be applied into the dataset to make it more suitable and convenient for the concerning algorithm models to perform well. This process in more challenging and should be done with utter most care if model accuracy is highly valued. Some of the processes that can be executed in this stage upon the dataset would be,

- Deletion of records having null values as a majority.
- Deletion of the columns which are irrelevant for recommendation purposes.
- Processing and converting value types to match with the algorithm types being considered.
- Removal of car records which do not depict the price values.
- Applying weights for attributes that contribute the most in recommendation.

3.5 Model Implementation

In the proposed system, for the recommendation purpose, both TF-IDF (term frequency-inverse document frequency) and cosine similarity techniques, is advocated as an ensemble model. TF-IDF is primarily for text and description analyzation. It helps to simply measure the relevance of certain words in a document or a set of documents. Since the main focus is examining content-based algorithms furthermore, TF-IDF functionality and uses becomes more appropriate. In order to calculate a certain word's TF-IDF two different metrics are needed at first. They are namely, the term frequency, which can be calculated in the simplest way as getting the raw count of that particular word and the inverse document frequency of a word, throughout a set of documents, which basically provides the data on if the word is rare within the document or common. If the inverse document frequency is closer to 0, the more common it is. The nearer the value is to 0 the more common it is, or else the closer to 1, the less occurring it is. In the end, the two metrics are to be multiplied and a TF-IDF score is generated. The higher the score, the more pertinent the word in that considered classified advertisement. As discussed, this relevancy calculating procedure can be applied within this proposed system of recommendation as well.

The next metric, which is known as cosine similarity, can be used to know how similar two documents are, irrespective of their size comparisons. In a multi-dimensional space, it calculates the cosine angle between two vectors. In the vectors, it is arrays containing words of more than one document. When plotted as a multi-dimensional space, where a dimension can be seen as a word in a document, this cosine similarity apprehends, not the magnitude of the documents, but the angle. Nevertheless, by using Euclidean distance, the magnitude can be computed. The main advantageous factor of cosine similarity is if though two documents are being apart through a long Euclidean distance due to the size matter, those two could yet possess a closer angle between each other. For example, it is a scenario similar to the word 'Running' is mentioned 60 times in one document, and 20 in another document. The miniature the angle is between the two documents, the more similar they are. By incorporating both the TF-IDF and cosine similarity, an ensemble model is to be created to make an optimized recommendation model, making use of each other's strengths and advance the process.

3.6 Application Integration

The final developed recommendation model will be thoroughly tested for bugs and if present, those will be rectified furthermore. The integration with the JavaScript based 'Tievs' application's front and backend environments will be done after complete model verification. The model will be within the recommendation API, which is built using the flask python framework, along with the other solutions developed by the research group members. The content management system and the recommendation API will be resting within the same machine environment in order to support communication channels.

3.7 Project Requirements

3.7.1 Functional requirements

- User should be able to login to the system successfully.
- User should be able to access the classified advertisements without impediments.
- Signals should be sent to the recommendation API about the customer preferred classified from the client system.
- Recommendation system should select the top recommendation list from other overall recommendations.
- Rediscover new classifieds by the recommendation system, each time they are added into the system.

3.7.2 Non-Functional requirements

- Scalability
- Reliability
- Accuracy
- Speed
- Optimized performance

3.8 Gantt Chart

The following figure demonstrates the time schedule estimated for the project tasks.

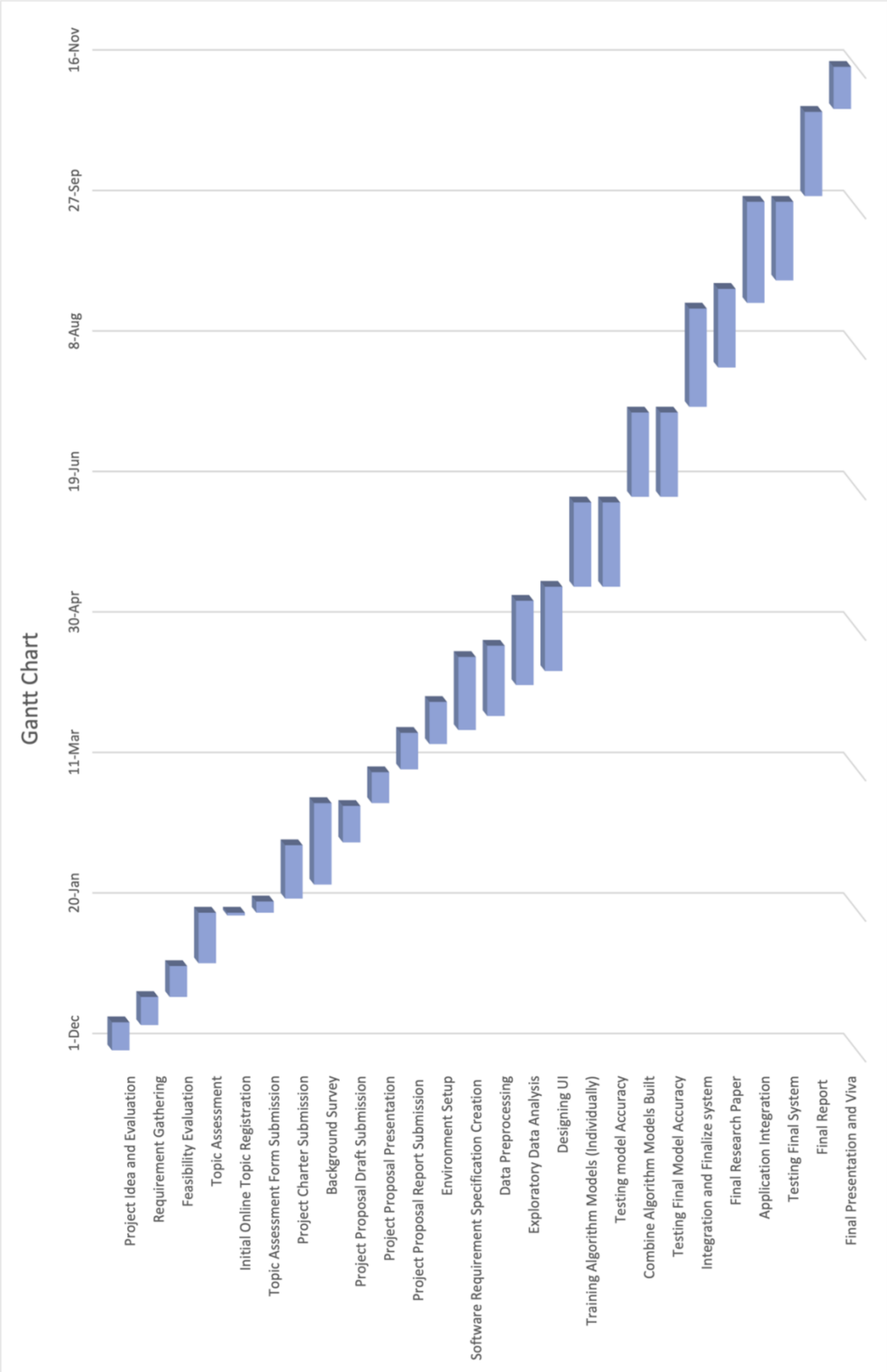

Figure 3.4 – Gantt Chart

4 DESCRIPTION OF PERSONAL AND FACILITIES

Customer-specific, optimized recommendation system

- Investigate suitable machine learning algorithm types available with regard to recommendation systems.

- Develop and train algorithm models encountered through investigation by utilizing the ‘Used Cars Dataset’ from Kaggle, containing necessary item characteristic details. (ex: Cars)

- Analyzation of the results of all eligible models and finalizing the most optimal model (most probably would be a hybrid model using the techniques of TF-IDF and cosine similarity), considering attributes such as speed, latency, scalability, accuracy etc.

- Implementation of the process for monitoring the customer behavior and gathering statistics, to forward into the selected model for further analysis and prediction of most favored and expectable item sets.

- Build relevant client and server-side components to visualize the analyzed and predicted, customer-specific search results and retrain the model using new information acquired from the application database.

5 BUDGET AND BUDGET JUSTIFICATION

Table 5.1 - Budget

Component	Amount (USD)	Amount (LKR)
1GB Memory – 1vCPU Droplet (Frontend)	60	12000
2GB Memory – 1vCPU Droplet (Backend & DB)	120	24000
.com Domain Name	12	2400
Total	**192**	**38400**

REFERENCE LIST

[1] B. P. P. C. Limited, “Kaidee to charge for some ads,” *Bangkok Post*.

[2] D. Das, L. Sahoo, and S. Datta, “A Survey on Recommendation System,” *Int. J. Comput. Appl.*, vol. 160, pp. 6–10, 2017, doi: 10.5120/ijca2017913081.

[3] R. Singla, S. Gupta, A. Gupta, and D. K. Vishwakarma, “FLEX: A Content Based Movie Recommender,” in *2020 International Conference for Emerging Technology (INCET)*, 2020, pp. 1–4, doi: 10.1109/INCET49848.2020.9154163.

[4] A. Pal, P. Parhi, and M. Aggarwal, “An improved content based collaborative filtering algorithm for movie recommendations,” in *2017 Tenth International Conference on Contemporary Computing (IC3)*, 2017, pp. 1–3, doi: 10.1109/IC3.2017.8284357.

[5] B. Kostek, “Listening to Live Music: Life Beyond Music Recommendation Systems,” in *2018 Joint Conference - Acoustics*, 2018, pp. 1–5, doi: 10.1109/ACOUSTICS.2018.8502385.

[6] S. Shaikh, S. Rathi and P. Janrao, "Recommendation System in E-Commerce Websites: A Graph Based Approached," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 2017, pp. 931-934, doi: 10.1109/IACC.2017.0189.

[7] G. Prabowol, M. Nasrun and R. A. Nugrahaeni, "Recommendations for Car Selection System Using Item-Based Collaborative Filtering (CF)," 2019 IEEE International Conference on Signals and Systems (ICSigSys), Bandung, Indonesia, 2019, pp. 116-119, doi: 10.1109/ICSIGSYS.2019.8811083.

[8] F. Sun, Y. Shi and W. Wang, "Content-Based Recommendation System Based on Vague Sets," 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 2013, pp. 294-297, doi: 10.1109/IHMSC.2013.218.

[9] R. Wita, K. Bubphachuen and J. Chawachat, "Content-Based Filtering Recommendation in Abstract Search Using Neo4j," 2017 21st International Computer Science and Engineering Conference (ICSEC), Bangkok, Thailand, 2017, pp. 1-5, doi: 10.1109/ICSEC.2017.8443957.

[10] H. Xue and D. Zhang, "A Recommendation Model Based on Content and Social Network," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 477-481, doi: 10.1109/ITAIC.2019.8785729.

[11] H. Chen, Y. Wu, M. Hor and C. Tang, "Fully content-based movie recommender system with feature extraction using neural network," 2017 International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 2017, pp. 504-509, doi: 10.1109/ICMLC.2017.8108968.

[12] “Most 5 Valuable Benefits of Agile Methodology | Blog – Denysys Corporation.” https://www.denysys.com/blog/5-benefits-of-agile-methodology/ (accessed Feb. 25, 2021).

[13] M. Alodadi and V. P. Janeja, "Similarity in Patient Support Forums Using TF-IDF and Cosine Similarity Metrics," 2015 International Conference on Healthcare Informatics, Dallas, TX, USA, 2015, pp. 521-522, doi: 10.1109/ICHI.2015.99.

APPENDIX

Plagiarism report