Temporary ID: TMP-21-068

Sri Lanka Institute of Information Technology

Project Topic Assessment – 2021

Topic

Tievs – A SMART online web application that enhances classified advertising through machine learning appliances.

Abstract (200 Words Max):

Online classified advertising platforms have been increasing popular in recent years, especially since most people do not utilize tangible information sources such as newspapers, magazines, booklets, Leaflets, etc. than in previous times, to post or refer to different advertisements regarding vehicles, houses, electronic items, pets, clothes, etc. with the intentions of selling or buying. Nevertheless, most platforms out there have not been in the appurtenant caliber in terms of providing a broad scope of customer types with user-friendly advertisement search options that are specialized according to their requirements all the while making that process more accessible and feasible or allowing core platform operations to be more self-learning, prognosticating and error-preventive. Hence it has become the goal of our 'Tievs' online web application to provide such facilities, all the while enhancing them to outsmart other products with similar intentions. Our customer-optimized, advanced recommendation system will suggest ideal advertisement search results specific to the interests and requirements of each customer, which then saves time and reduces the number of searches they have to execute manually on the web application, resulting in less turnover rate. When customers declare the price of the product before posting an advertisement, the system automatically generates and suggests a price amount considering the current market prices of similar characteristics included products. Besides, if any fraudulent advertisements were to be posted, for example, having fake/irrelevant images or irrelevant/inappropriate information contexts, the system will promptly identify and prevent such events from happening. As implied, Tievs will operate innovatively, to move forward in leaps and bounds to uplift online classified advertising to the next level.

Research Area/Group: Select the area by referring to the document uploaded to the Courseweb

ICT for Development

Supervisor should fill this part

Supervisor: I certify here that co-supervisor and myself can guide this students to acquire required knowledge skills and attitudes pertaining to above sub domains of his/her specialization.

Supervisor: **Ms.Manori Gamage**

Signature

Continuation of Previous Year Project? ☐

If yes, state the Project ID

and year

Co-Supervisor: **Ms. P. K. Suriya Kumari**

Signature

External Supervisor

Name

Team Members:

Student Name	Student ID	Specialization
Leader: R.A.D. Prathapa	IT18122060	IT
Member 2: K. S. S. Bandaranayake	IT18113532	IT
Member 3: J.M.S. Ravihari	IT18089400	IT
Member 4: S.K.A.K.I Madhushani	IT18082548	IT

Research Problem:

Even though it is distinctly apparent that online classified advertising platforms have already been implemented by numerous parties in recent years, there is yet to be a highly ameliorated application that incorporates complex machine learning technology to enhance the prominent functionalities to facilitate customer expectations and requirements efficiently. It is undoubtedly troublesome when people have to interact with an application that is not much user-centric and accommodating towards their needs, (like most available applications nowadays) considering their busy schedules. Most applications provide the customers only with similar items for the current item search but they would not be optimized for particular customer needs specifically, considering into depth characteristics of the searching product such as color type, manufactured year, price ranges, etc. Furthermore, having less scalable, less accurate, high latency processes running internally for recommendation of products, can degrade the applications' overall productive performance. In addition, there can be instances where one can preconfigure the necessary characteristics prior to searching the relevant product, although those appliances can be particularly difficult to be utilized by elderly user groups and non-technical individuals, who can be defined as a large part of the target audience of classified advertising. Hence it will evidently cause a waste of time in finding the desired result, leading to a high customer turnover rate as well [1].

For customers who utilize those platforms to post their relevant advertisements targeting other customers, it can be strenuous when clarifying the accurate price for the product to be sold. Then those customers will be doing exhaustive research themselves, to find out what is the exact selling price of other similar level and type products on sale, which takes a huge amount of time out of their schedule again and as a whole, reduce user satisfaction level and comfortableness provided by the application[2]. User-friendly classified advertising platforms that provide the price predictions for selling products are obscure. In addition, from the applications' developers and owner's point of view, they have to implement strict program logic to identify or prevent fraudulent advertisements being posted (fake images/ irrelevant descriptions) accidentally or deliberately by users of the system, through verification and validation processes internally. This will require to inspect posts separately and meticulously, even after the ad being posted to the customers and if so, customers will view those fraud ads and lose confidence in using the platform for their needs, enabling loss of application reputation as well [3]. Most classified advertising products currently in use do not prioritize detecting and preventing fake advertisement images or characteristic details from being submitted, let alone inspect them internally and inform the respected users who are in the process of submitting the advertisement. Even more, many implemented online classified advertising systems, simply do not exhibit rich user interfaces for smooth functionality or promote quality user experience and user friendliness, in consonance with latest trending technologies.

References

[1]   A. A. Patel and J. N. Dharwa, "An integrated hybrid recommendation model using graph database," 2016 International Conference on ICT in Business Industry & Government (ICTBIG), Indore, pp. 1-5, 2016, doi: 10.1109/ICTBIG.2016.7892680.

[2]   Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques," *International Journal of Information and Computation Technology,* vol. 4, no. 7, pp. 753-764, 2014 .

[3]   Maktabar, Mahdi and Zainal, Anazida and Maarof, Mohd Aizaini and Kassim, Mohamad Nizam, "Content based fraudulent website detection using supervised machine learning techniques in International Conference on Health Information Science," Springer, 2017.

Solution proposed:

A SMART web application for classified advertising called 'Tievs' will be built with the use of machine learning algorithms to cater the requirements of customers who visit the application for both advertisement viewing and posting purposes, having the following features.

- Customers who are visiting 'Tievs' to view and search for different items on sale, will experience an optimized, recommending , powered by machine learning appliances, that will suggest the most preferred item range for that particular customer, based on the provided characteristic details of the demanded item type. The approach taken to achieve this will be done by comparing various machine learning algorithms and opting for the most convenient one, considering parameters such as accuracy, speed, scalability, latency, and application wise effectiveness. The initial input data for training the selected algorithm will be taken from an external source having the required dataset. After application launch, it will be visible that this approach would be more user-friendly and enriched overall, with comparison to already implemented recommendation systems.

- When customers with the purpose of posting an advertisement on behalf of selling a product of theirs's, visit the application, they would want to know about the most convenient price to apply to the post and Tievs will make their task easy by providing an estimated, predicted dynamic price, considering the current selling prices of similar products in market, prior to the advertisement getting submitted into the system. This will be providing customer satisfaction and efficiency of the overall process since the customer themselves won't be needing to do private research on the accurate price. The most optimal machine learning algorithm will be chosen by doing further research and inspections as to which gives the fastest and most precise result for the predicted price and the process will include getting an appropriate and sufficient dataset, from an external source, to train the eligible models. Through that model, the price will be predicated for the given characteristics of the concerning item type.

- Mostly, spam is identified only after it occurs. Actions are then implemented to prevent it from happening again. 'Tives' will prevent posting the irrelevant content to the system before the significant damage is done. Initially this will be achieved by considering different algorithms for the most appropriate one and Implementing the supervised learned model to detect the advertisement being inspected by inappropriate content. Required data set will be gathered by external source and prior to posting an advertisement, the application will inspect for any spam/irrelevant information contexts regarding the specific item currently getting posted. Finally, customer will be prompted with a warning message If the advertisement is being inspected by spamming. This will be increased the system more realistic and provides the user satisfaction.

- Fake or irrelevant images must be identified before a user post an advert containing one, rather than examining for such images after it has been submitted, if it's essential to handle future conflicts. Through training all competent image processing algorithm models with predefined image data, which targets a specific item range, (Ex: Vehicles) and identifying the most optimal image processing algorithm model that, additionally, supports 'Tievs' requirements, the application will be able to successfully inspect whether there are any fake/inappropriate images, that are going to be uploaded with the advertisement form by the customers.

System Overview Diagram for the solution proposed. Recommended to draw using draw.io. Note: This is not an activity/flow (UML) diagram

1. **Man components including the data sources, stakeholders, interaction among the stakeholders, etc.**
2. **Interconnection among the components**
3. **Major SW and HW components**
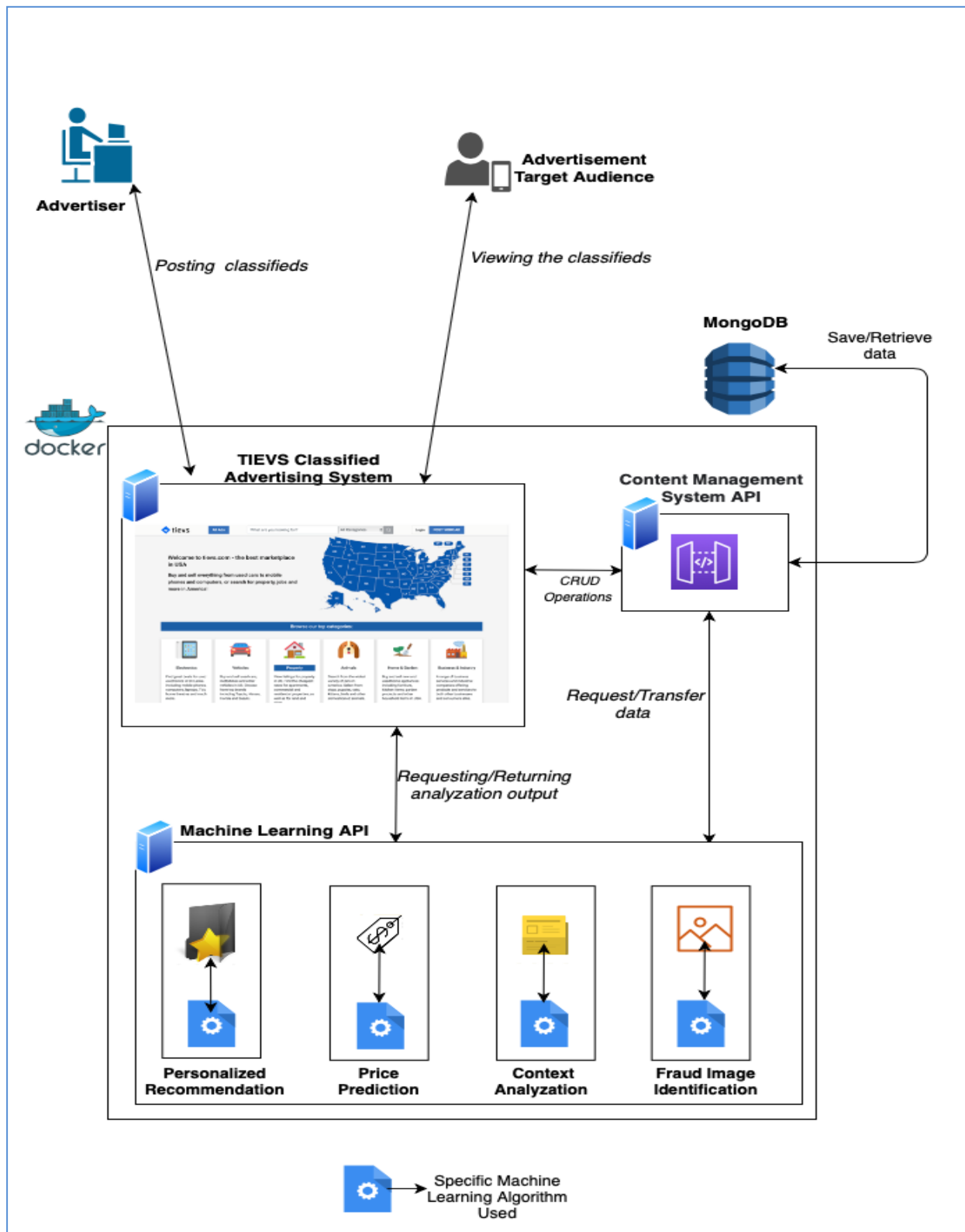
Objectives (1 main objective and 4 sub objectives):

Main Objective: Implementing a SMART, systematic, and rich user-friendly online classified advertising platform, to surpass its predecessor systems in delivering and catering to customer's necessities, using the most optimal, cost-effective, timesaving, human-centric and error preventive approaches.

Sub Objective 1: Developing an optimized item recommendation process, using a convenient machine learning model, to provide personalized content to customers in an interactive manner.

Sub Objective 2: Implementing a dynamic price predicting system to recommend a selling price value for an item of a specific type (such as Cars), using a predictive machine learning algorithm.

Sub Objective 3: Proactively identifying and preventing fraudulent advertisement content information, from being posted in the application, using machine learning appliances.

Sub Objective 4: Detection of fake, ill-matched, conflicting or unsuitable images within the advertisement form, and displaying cautious warnings, using image processing techniques.

Task divided among the members

(For the research purpose, we will be targeting one type of classified such as Cars)

Member 1 : 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 (might even be a hybrid model), 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.

Member 2 : Dynamic price prediction and suggestion

- Analyze price impactful item features and investigate for an efficient predictive algorithm model capable of forecasting a selling price for an item. (considering accuracy, efficiency, speed, effectiveness, size of training data, number of features productively supported, etc.)

- Develop and train the models, with refinements of parameters, using the preprocessed version of the 'Used Cars Dataset' acquired from Kaggle, and realize the most competent model of all. (can be a single type or a hybrid version of multiple types)

- Build and develop necessary client side and server-side components for data extraction of an item from advertisement form prior to posting.

- Integration with developed algorithm model and visualization of the predicted price for the customers in form of a suggestion.

- Continuous retraining of the model through a renewed dataset, extracted from current item information within the application itself since some customers submit their own price values, rather than the system suggest price, and those are necessary for future prediction processes as well.

Member 3 : Proactive identification and avoidance of fraudulent advertisement information context

- Explore the most appropriate machine learning algorithm by considering the Accuracy and / Interpretability of the output Speed or Training time, Linearity etc., to identify fraudulent advertisement information context having inappropriate/irrelevant content.

- Congregate the relevant data set using external sources (e.g. Kaggle - car_ad.csv/ vehicles.csv) or survey using craigslist region to train the model and test the model built.

- Develop and Train the model to detect, whether the advertisement being inspected by inappropriate /irrelevant content.

- Build the appropriate client and server-side components and extract the data from advertisement form, prior to posting.

- If the advertisement is consisting of the inappropriate / irrelevant content, the client will be prompted with a warning message, indicating such issues present in the advertisement, which is yet to be submitted.

Member 4 : Detection and prevention of conflicting, unsuitable images in the advertisement form

- Identification of the most efficient image processing mechanism, deriving the best results promptly for the Tievs application, through investigations done on many other available competitive image processing algorithms.

- Selection of the best performing algorithm for the application will be done. Images will be givens as input (Ex: Car interior/exterior images) with their respective identifying labels to train and develop an acceptable model.

- Build and develop required client and server-side app functionalities and interfaces for integration with the algorithm instantiated.

- Monitor and analyzation of the images attached in the advertisement form to find its correlation to the relevant item type, using the developed model.

- Visualize cautious notifications for customers to discourage proceeding further with the submission if any conflicts occur.

Technologies to be used:

Machine learning frameworks and algorithms (for recommendation and predictions)

Image Processing

Angular Framework

Python

Node.js

Strapi

MongoDB

Docker

If supervisor States that this year is a continuation of previous work, state the further work the students should do compared to the previous years.
<mark>(NOTE: This part has to be filled by the supervisor)</mark>

N/A

<u>This part will be filled by the Topic Screening Panel members</u>

Acceptable: Mark/select as necessary

Acceptance/ Rejection	Correction State	
	Minor Correction	Major Corrections
Accepted	☐	☐
Resubmit	☒	☐
Rejected	☐	

Corrections (if necessary)

The proposal is not focussed overall

Explanations should be more concise. For example, objectives should be smart, actionable goals which are maximum 1-2 sentences each.

When explaining, simply focus on the "what" and "how" in this format e.g.,
"I will do X, using Y, resulting in Z", in just enough detail. The readership of these proposals are educated computer scientists, not the general public, therefore explaining things from first principles upwards is not required.

Datasets are required to achieve certain functionalities promised, please indicate how you will obtain those datasets. E.g., "system would provide the predicted dynamic price"

Any other Comments:

_____good_____

_____

_____

____

Approved by the review panel:

Panel 01:

Member's Name	Signature
Ms.Pradeepa Senani Bandara	
Ms. Bhagyanie Chathurika	

Panel 02:

Member's Name	Signature
Mr. Jayantha Amararachchi	
Dr. Shyam Reyal	
Mr.Prasanna Sumathipala	

Important:

1. According to the comments given by the panel, do the necessary modifications and get the approval by the **same panel**.
2. If the project topic is rejected, find out a new topic and inform the CDAP Group for a new topic pre-assessment.
3. A form approved by the panel must be attached to the **Project Charter Form**.

Appendices

Appendix 1 – Endorsement response by supervisor Ms. Manori Gamage

2021-Reg-Topic Assessment form of TMP-21-068 for Supervisor Ms. Manori Gamage Endorsement"

Manori Gamage <manori.g@sliit.lk>
Fri 22/01/2021 12:50
To: K.S.S Bandaranayake it18113532
Cc: Prathapa R. A. D. it18122060; Ravihari J.M.S. it18089400; MADHUSHANI S.K.A.K.I. it18082548; CDAP SLIIT

checked

Manori
Gamage

Senior Lecturer
SLIIT | Metro Campus

. . .

Reply | Reply all | Forward

K.S.S Bandaranayake it18113532
Fri 22/01/2021 10:15
To: Manori Gamage
Cc: Prathapa R. A. D. it18122060; Ravihari J.M.S. it18089400; MADHUSHANI S.K.A.K.I. it18082548

TopicAssessment_1_TMP-21-...
420 KB

Dear Madam,

Please find the attached Topic Assessment form for endorsement herewith.

Best Regards,
IT18113532
K S S Bandaranayake

Appendix 2 - Endorsement response by co-supervisor Ms. Suriyaa Kumari

2021-Reg-Topic Assessment form of TMP-21-068 for Co-Supervisor Ms. Suriyaa Kumari Endorsement   📎 1 ∨ ⊞

Suriyaa Kumari <suriyaa.k@sliit.lk>   👍 ↩ ↞ → ⋯
Fri 22/01/2021 10:20
To: K.S.S Bandaranayake it18113532
Cc: Prathapa R. A. D. it18122060; Ravihari J.M.S. it18089400; MADHUSHANI S.K.A.K.I. it18082548; CDAP SLIIT

Dear Students

I confirm the topic assessment form submission

Best Regards

Ms.Suri┌─────────────────────────────┐
       │ Thank you for your confirmation. │
       └─────────────────────────────┘

Get Outlook for Android

⋯

| Thank you! | **Thank you for your confirmation.** | Great, thanks! |

💬 Are the suggestions above helpful?   Yes   No

Reply | **Reply all** | **Forward**

KI ✓ K.S.S Bandaranayake it18113532   📎
Dear Madam, Please find the attached Topic Assessment form for endorsement herewith. Best Regards, IT1...   Fri 22/01/2021 10:18