Hooman – An Online Dating Website

Jayasekara K.R.T.
Department of Information Technology SLIIT
Malabe, Sri Lanka

Manori Gamage
Department of Information Technology SLIIT
Malabe, Sri Lanka

Ginigaddare G.D.U.D.
Department of Information Technology SLIIT
Malabe, Sri Lanka

Hansika Mahadikara
Department of Information Technology SLIIT
Malabe, Sri Lanka

Wijerathne M.K.N.B
Department of Information Technology SLIIT
Malabe, Sri Lanka

Senarathna A.M.M.D.N
Department of Information Technology SLIIT
Malabe, Sri Lanka

Abstract:- In society there are two types of people some are introverts and extroverts. Introverts prefer not to be socialized much and could be shy to talk to people, but in the end, we areall humans we all need someone to be with not to be lonely, soat a point in their life they need a mate to live with, so theycould date or marry someone, some of the introverts don't know how to start a conversation even if they do most of the time they would feel awkward when having a conversation, to overcome this new advanced series of features have been added. Using a profile matcher with a series of questions shows what would be their personality and by grading them, users who have the same grades will be matched, this would make it easier for peopleto get connected with one another because they might have the same hobbies and same interests, so the bonding experience is easy, as an advanced feature, having a horoscope matcher isvery valuable for the people in Asia, especially in Sri Lanka because of the culture. Here you need to give the image of the horoscope and it automatically matches with others and gives suggestions, an intelligent message supporter as a feature, some introverts have no idea what they have to say if another personis texting them, so using this intelligent message supporter we suggest what are the responses that user can give to the other person. Video games tend to attract more people just because they are interactive and fun and eventful, so a video game is perfect to predict the strengths and weaknesses of the player in order to give the player's partner an idea about them.

Keywords:- Introverts, Profile Matcher, Horoscope Matcher, Intelligent Message Supporter, Video Game.

I. INTRODUCTION

When it comes to online dating, the person we like can be a psychopath or worse even though from the first impressionto be the most honest and innocent person. But if we canget to know this person to some extent the risks can be mitigated, so using some advanced features with dating sites will help users to be more careful and to pick a decent person to spend the rest of their life, such as having a personalitytest using an algorithm, the system calculates the personalityof the user and suggests mutual users same as the user'spersonality, but this is not enough if the users are more maturedwith the system and found a good match for them their next step would be the marriage or some of the people start their relationship by matching horoscopes this is mostly happening in Sri Lanka because we have a culture to maintain and most family members approve the partner only if the horoscope getsmatched. So as our second feature we can use a horoscope matching feature, but this is not enough to get to know one another because the horoscope doesn't say anything aboutwhat kind of a person they are, but what if they could know what the strengths and weaknesses are of each other in aromantic relationship, this way they can get an idea about eachother. A game will analyze the strengths and weaknesses ofthe player. Finally, an intelligent chat service most introverts need to be trained before talking to people because most of them don't know how to respond to other people's texts sothe intelligent message supporter gives them suggestions fromprevious chats, if a similar conversation happened before then this chatbot gives what to say, so you can select it and sendit. Regarding the game, two game levels have been developed where each level predicts a strength or a weakness. In level one, the listening skill of the player will be predicted. In level one player must navigate through a maze with the help of audio clips [6][7]. There are invisible score points inside the maze if the player goes along the correct path according to theaudio clips score points will be added to the score and there arepenalty points that will be used to deduct the score points of the player if they went off course. A Machine Learning model will help to predict the strengths and weaknesses of the player.The second level is somewhat trickier because it measures problem-solving skills. It's a puzzle-solving game and this level will improve and predict the user's problemsolving skills. According to Mayer and Wittrock (1996, 2006) [5], there area few characteristics of problem-solving.

1) It is a cognitive process.
2) It is goal oriented.
3) Complexity (and hence difficulty) of the problem depends on one's current knowledge and skills. and since

puzzle games have these characteristics and cantap into multiple cognitive abilities [8] puzzle games can be used to measure and improve problem-solving skills of a player.

II. LITERATURE REVIEW

Over time there are many features came up with dating platforms and the community also talked about developing some advanced features, but these were harder to do when we are talking about Tinder, they just give photos of people to swipe left or right if they like them. This was not an accurate approach to matching someone. When it comes to online dating there is a saying, "you have to kiss a lot of frogs to find a prince". Meeting someone online is fundamentally different than meeting someone in real life [1]. Therefore, having a personality test describes more about the user to the system then using machine learning we can suggest what are the best matches for them so the user can see only people who have similar personalities just like them which makes it more exciting. If the user trusts the system, then the user will not be afraid and move on with the date. When it comes to matching people with other people within Sri Lanka or within South Asia horoscope can play a main role but this feature is not implemented on any dating website because people had not thought to a greater extent than people have proposal matching websites[5] but you need to have a partner to match your proposal this is always not easy so if we use those implementations in a dating website people can use their horoscopes and match it with the person who they got suggested with if it is and if the user likes the matched person, then they can even build up a future without hesitating.This horoscope matching is included in our system, and itis consulted and instructed to be made by a real astrologer because most of the time online horoscope matching could be wrong and end up in failed marriages [2].

III. METHODOLOGY

To implement this we used multiple number of data sets because each of these systems we different from each other and only two of them need to have data sets that would be the intelligent chat bot and "profile suggesting" components, both these data sets were gained from Kaggle, one dataset is used to analyze the conversations of people in movies and if the models were trained with them then model learns howto think like a normal human, the main purpose is to getinformation about the user and suggest the best match possible for them therefore when it comes to horoscope matching all need to be researched from many articles and find the patternsin them and understand how the concept work this was done by learning from astrologer to match the patterns and this" information " is made by the researcher himself and two array sets have been used to train the gaming model to get the relationship between the score of the game(1st array set - 5,10, 15, ...) and a value that need to be given (2nd array set -0, 0, 1, 1,...) which is between 0 and 1.
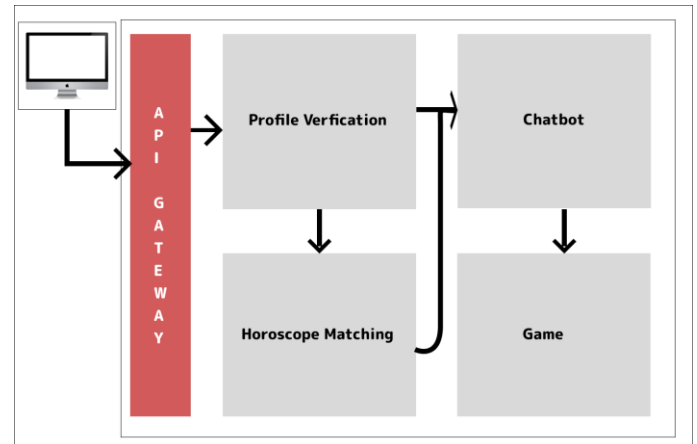

Fig.1. overall diagram

Intelligent Message Supporter. In the conversational dataset, responses are grouped as sender and receiver removing the stop words and making a bag of words so the unique and meaningful words are in that bag so the response is generated from this bag of words, definitely, these sentences are not going to make any sense because they cannot be grammaticallycorrect since there are no stop words and might not make any sense either, because processing sentences, so while we are doing feature engineering, we make two arrays to store these sender and responders values after that using seq2seq model and LSTM[1] we make a relationship between these two groups and categorical cross entropy as the loss function, therefore, it will calculate how bad the guess is and give itto the optimizer Adam so it can make a better prediction asfor the next prediction. This is the machine learning model training process, now this model is stored in the frontend and when the user types in the chat the text passes into the Machine Learning model and predicts what would be the response, and using the tokenization method, the system tokenizes the text, so it has a numerical value so it's easier to find the similar sentences which have the same words and token values as the response, this can be a range of words and limited it to asthree sentences these stored responses are used by the users who used the chat earlier. Seq2seq problems are a special class of Sequence modeling problems, both these problems are regarding input and output of sequences. Specially Encoder- Decoder models were used to solve Seq2Seq problems. For the encoder we pass the input sequence according to this scenario we pass the text which user types as the message, from the encoder it generates a vector, this is a collection of tokens and this is always fixed length this is called as 'context vector' thisheld the whole meaning of the sentence as a vector, this vectoris making the decoder easier to make the accurate predictions. The decoder reads this vector and goes from each token to token and targets the sequence token by token [2].
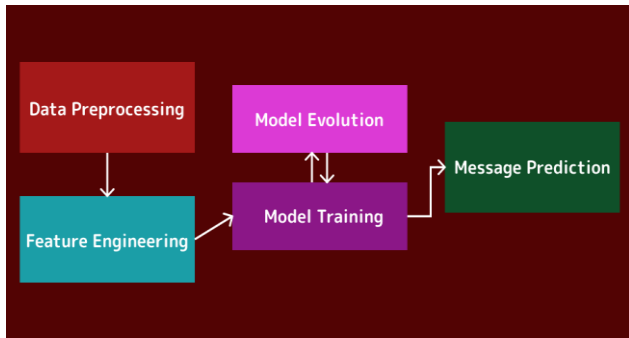
Fig.2. Basic process structure in Intelligent Message Supporter (Chatbot)

Fig.3. Summary of the model

Profile Verification System. A Convolutional Neural net- work, which is one of the main image processing modelsused in neural networks has been trained in identifying char- acteristics of images related to face attributes. As per image preprocessing, all the images are converted to grayscale to reduce the noise of the images. The following classifications were conducted by processing images using CNN's. Four models have been built using CNN in order to classify age, gender, hair, and nose. Based on the model and its classifica- tion accuracy, convolution layers were used with an activation function. The dense layer of the output layer has a softmax activation function according to categories that needs to be classified by the model.

Fig.4. Model Training in the Profile Verification System

A training test split is used to sample the training dataset against the test data set. Model checkpoint has been used tosave the best model with the highest accuracies.

Horoscope Matching System Users must upload an im- age of their horoscope. Using image processing, The

main technologies that are used to build this component are image processing and CNN (Convolutional Natural Network) which are used for building deep learning models. This CNN behaves like a human brain when it is identifying images [10],[11]. When we use this CNN, it needs a data set to train the Machine Learning model to identify the combinations betweenhoroscopes since there were no found data for this purpose to be found on the internet this data was made by the researcher himself with the help of a professional Astrologer. There are 12 astrological signs and this astrological sign when we are writing in Sinhala it can be different in many ways it changes from person to person therefore only from one astrological sign researcher took 250 to 300 images to train what is the written name looks like to train the model to capture the imageusing CNN so that model knows exactly what the astrological sign is accurately. So, it has a collective knowledge of around more than 3000 images, so the model gets the value correct way [14]. Now for the Machine Learning algorithm this component passes the Astrological sign that's detected anddo the prediction using this method matching the horoscopeis promising and the results are very accurate also since this needs to improve much more later as for the start this is avery good start and a very good method to math horoscopes, the accuracy of the Machine Learning model is high more like85as shown in the Fig 4), when considering accuracy this isa good rate
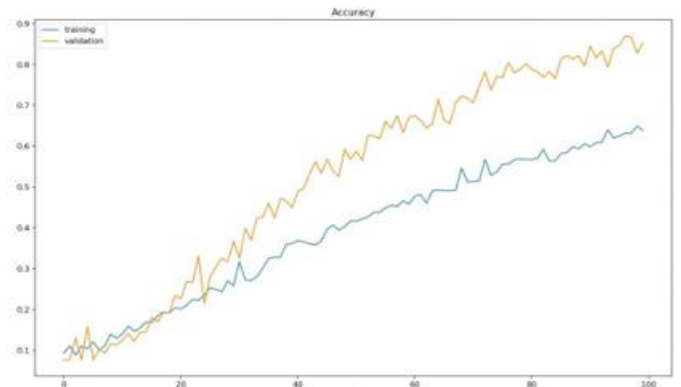

Fig.5. Model training Diagram

Fig.6. Model training in Horoscope Matching Component

After training is finished Machine Learning model needs to be deployed to the backend so from the frontend imageis passed to the backend of our system as a POST

request through the API. Earlier CNN was used to process the data of the image and the accuracy of the image detection was high from that but before moving on to the Machine Learning model Optical Character Recognition (OCR) is used to detect the image's data, to process the image and get the data. To do this we have used 550 images to train the model, but the accuracy of the result is very poor (accuracy of 34 Because of that we moved on to CNN, OCR process is done by the predefined model itself, but this CNN model is made by the researcher himself.

Strength and Weakness finder (Game) Mainly used lo- gistic regression to implement this component. To reduce code complexity (Unity accesses the ONNX model as an asset so there is no need for hard coding in unity). Especially in a videogame, we should maintain less code complexity because it will cause GPU CPU to be in high usage and video games use GPU and CPU heavily so code complexity will affect GPU and CPU as a result it will affect the game performance as the game grows and becomes more complex with the future works. We use previously trained data (an array set of the score given by each level and an array for results according to the score) which will lead to more accurate results (accuracy- 0.89) than not using Machine Learning. For example, if we use if-else we will only know if the player has good listening skills or poor listening skills. If we want to get the percentage of the listening skill, we have to write more code which will lead to more complexity of the game (as mentioned above). It is better to use a Machine Learning algorithm to predict the player's strengths and weaknesses which was trained using training data to get a more accurate result with the percentage of the listening skill because even if the player got the full score, it doesn't mean that the player has the best listening skill so prediction using previously trained data is the best option here. There are two levels in this game for the second level we mainly focus on how good the user is when it comes to solving problems, for this purpose a jigsaw puzzle is used, When the player solves the puzzle and if they complete the puzzle with extra time left on the timer, they get extra points but if they take longer to complete the puzzle than the given time a penalty will impose. So, the result that goes into the system for the Machine Learning model will be reduced. And for every correct placement, the game score will be increased.

IV. RESULTS AND DISCUSSION

It's a highly advanced system and much harder to maneuver since we have to predict everything that user does when they logged into the system, there is always a probability the rate of being successful, and models need to be trained more, that's what we have found so far but the probability of the success rate is high since we use actual conversations of people and model prediction is most of the time suggests pretty much the correct reply that we need to use for but the accuracy is not much likable and need to develop a lot more than right now.
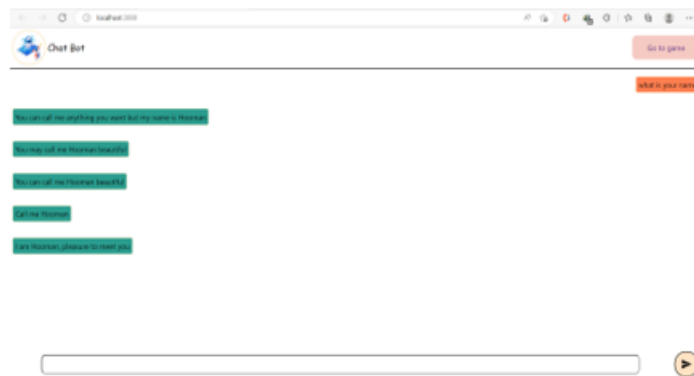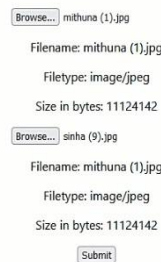

Fig.7. Chat screen of the Intelligent Message Supporter

Fig.8. Profile Verification System result interface
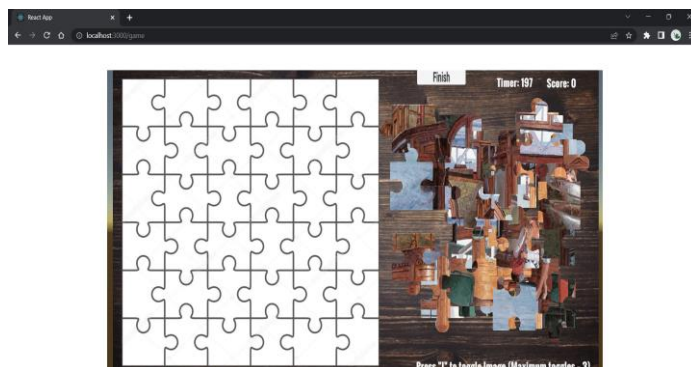

Fig.9. Horoscope matching result

Fig.10. Gamification interface

So, what we really need to work on more right now is not about the optimizers or the loss function but the data preprocessing part since this is a product that needs to be with people to help get their shyness and unfamiliarity about how to talk. The chatbot can help them a lot

V. CONCLUSION AND FUTURE WORK

Since this model is deployed with the website it will learn new words, new sentences and everything but that's not enough, need to monitor how regularly people use this model and the answers it suggests to them if they are using it more often then it's great but what if the suggestions are not suitableat all, so we have to monitor it, by collecting the data thatusers used to chat with each other, more like chats peopleare having so we can detect what are the areas that they are more commonly focused on and if that area needs to have an improvement focusing on that area can improve the chatbota lot. Always this needs to be more focused on the empathy side because all humans need to be treated nicely and gently. As for the game component, it can,

1) Use more levels to get more strengths or weaknesses of the player
2) We can use this method to get not only strengths and weaknesses but also personality traits too.

VI. ACKNOWLEDGMENT

REFERENCES

[1]. Y. Wan, "Short-Text Sentiment Classification Based on Graph- LSTM," 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 35-38, doi: 10.1109/AIAM48774.2019.00014.

[2]. L. Zhang and H. Deng, "Sentence Simplification Based on Multi-Stage Encoder Model," in IEEE Access, vol. 7, pp. 174248-174256, 2019, doi: 10.1109/ACCESS.2019.2957160.

[3]. S. Al-Rousan, A. Abuhussein, F. Alsubaei, O. Kahveci, H. Farra and S. Shiva, "Social-Guard: Detecting Scammers in Online Dating," 2020 IEEE International Conference on Electro Information Technology (EIT), 2020, pp. 416-422, doi: 10.1109/EIT48999.2020.9208268.

[4]. T. Ghosal, S. K. Das and S. Bhattacharjee, "Sentiment analysis on (Bengali horoscope) corpus," 2015 Annual IEEE India Conference (INDICON), 2015, pp. 1-6, doi: 10.1109/INDICON.2015.7443551.

[5]. P. Alexander, P. Winne, R. Mayer and M. Wittrock, Handbook of educational psychology, 2nd ed. Routledge Taylor and Francis Group, 2006, pp. 287-303..

[6]. N. Mead and D. Rubin, "Assessing Listening and Speaking Skills. ERIC Digest.", Ericdigests.org, 1985.

[7]. J. Peragine and G. Hudgins, 365 Low or No Cost Workplace Teambuild- ing Activities: Games and Exercised Designed to Build Trust Encourage Teamwork Among Employees, 2nd ed. Ocala, Florida: Atlantic Publish-ing Group Inc, 2016, p. 153.

[8]. P. Fissler, O. Küster, D. Laptinskaya, L. Loy, C. von Arnim and I. Kolassa, "Jigsaw Puzzling Taps Multiple Cognitive Abilities and Is a Potential Protective Factor for Cognitive Aging", Frontiers in Aging Neuroscience, vol. 10, 2018. Available: 10.3389/fnagi.2018.00299.

[9]. Q. Wang, L. Sun and Z. Chen, "Sentiment Analysis of Reviews Based on Deep Learning Model," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), 2019, pp. 258-261, doi: 10.1109/ICIS46139.2019.8940267.

[10]. Huiping Li, D. Doermann and O. Kia, "Automatic text detection and tracking in digital video," in IEEE Transactions on Image Processing, vol. 9, no. 1, pp. 147- 156, Jan. 2000, doi: 10.1109/83.817607.

[11]. Mattohti, M. Ibrahim, A. Dawut and A. Hamdulla, "Effective Text Detection in Composite Images with Simple Background," 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE), 2019, pp. 263-266, doi: 10.1109/ICIS- CAE48440.2019.221631.

[12]. P. Ranjitha, K. Rajashekar and Shamjith, "A Review on Text Detection from Multi-Oriented Text Images in Different Ap- proaches," 2020 International Conference on Electronics and Sus- tainable Communication Systems (ICESC), 2020, pp. 240-245, doi: 10.1109/ICESC48915.2020.9156002.

[13]. L. Rajamanthri and S. Thelijjagoda, "Plagiarism Detection Tool for Sinhala Language with Internet Resources Using Natural Language Processing," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 2021, pp. 156-160, doi: 10.1109/ICIAfS52090.2021.9605852.

[14]. S. Malakar and W. Chiracharit, "Thai Text Detection and Classification Using Convolutional Neural Network," 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2020, pp. 99-102, doi: 10.23919/SICE48898.2020.9240290.

[15]. W. He, X. -Y. Zhang, F. Yin and C. -L. Liu, "MultiOriented and Multi-Lingual Scene Text Detection With Direct Regression," in IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5406-5419, Nov. 2018, doi: 10.1109/TIP.2018.2855399.

[16]. V. Skorpil, V. Oujezsky, P. Cika and M. Tuleja, "Parallel Processing of Genetic Algorithms in Python Language," 2019 PhotonIcs Electromag- netics Research Symposium - Spring (PIERS-Spring), 2019, pp. 3727- 3731, doi: 10.1109/PIERS-Spring46901.2019.9017332.