For

“Was it true or false?”

By:

Abstract:

The increasing availability of eye-tracking technologies has provided researchers with new opportunities to gain deeper understanding of the interpretive process. These tools can also help them identify the factors that influence visual perception and make informed decisions. Using this technology, a classification model was developed to find out if the person has true or false memory about the event. The project was carried out through the use of an eye tracking device known as Tobii, which collected the data of 23 participants. Since it is a supervised learning and multi-class classification problem we used the popular classification algorithms like Logistic Regression, K-Nearest Neighbours, Decision Trees and Random Forest for the model creation. For evaluating the model performance, the accuracy score is calculated for each classifier. The Classification report of the most accurate model is evaluated in terms of precision, accuracy, recall and F1 Score. The accuracy was high for the RandomForest classifier with a value of 68% and different methods were tried in order to improve performance of the model.

Keywords: Eye tracking, True Memory, False Memory, Tobbi

Contents

1. Introduction:

Due to the increasing interest in the interactions between humans and computers, the field of artificial intelligence (AI) is becoming more prevalent. This is evidenced by the increasing number of studies being conducted on the development of novel AI systems that can perform various tasks at human-level. One of these is image classification, which can be performed automatically and effectively with minimal human intervention. We tend to forget about the things that we have learned and experienced because we have become so used to having all this information about ourselves. A memory is a kind of information that can be stored in an unconscious or conscious manner. This process can be described as the formation of a set of systems that can act in a cooperative manner.

Due to the complexity of the process, it can be prone to errors in the testimony that a hearing witness provides. A true memory is a type of information that can be retrieved from various sources. It can be anything that is related to a particular event, such as a visual or virtual image. People are prone to experiencing errors in their memory due to various factors. False memory is not a simple mistake, while it is common for people to mistakenly think that they have a memory error. It involves a level of certitude that the memory is still valid. Every person experiences a memory failure at some point. A false memory is different from a memory that is composed of facts that happened in the past. Instead of remembering the details of an event, a false memory focuses on the things that were never experienced.

During the 1990s, people who were going through therapy for various issues started developing false memories of violent and traumatic experiences. According to the therapists who worked with them, these individuals had been victims of childhood abuse. These memories were then produced by the individuals who were recovering from these experiences. According to psychologists, these types of memories can also be produced in real life. For instance, people might create false memories of talking about their friends on the street. However, instead of their friends' names, they talk about their acquaintances' names. A study revealed that people's memory confidence is related to their accuracy. They were more likely to retrieve false memories when they were presented with a specific theme. This suggests that people have a higher level of confidence when it comes to recalling accurate memories.

There is also evidence contradicting this idea that people with high confidence are more likely to retrieve false memories. For instance, people with high subjective confidence are more likely to report having a high accuracy when it comes to recalling false memories. When we think about a beach scene, most of the objects that we see are usually related to the theme of the scene, such as the sand, water, and sun loungers. However, the item "the beach ball" is not necessarily a part of the scene's memory. In 1995, a study revealed that people were more likely to identify objects with accuracy if presented with a false memory concept. To determine how people perform certain tasks related to retrieval, the researchers conducted a study that involved monitoring the eye movements of participants. They then used these data to measure the accuracy of their actions. They were able to determine how these movements affected their accuracy. Eye-tracking is a process that involves the measurement and tracking of the user's movements and focus point. It is commonly used in various fields such as marketing, psychology, and computer gaming. Due to its applications in cognitive science and computer science, eye tracking is becoming more prevalent. Through the use of a camera or an eye-tracking sensor, users can easily collect and analyse data related to their movements. This technology is very useful and can be widely used in the future. In addition to being used for classification tasks, it can also be used for various other tasks.

The paper is divided into four sections, the first section includes a brief introduction followed by a study based on previous papers. The next section describes the methodology of the project, which includes the detailed explanation of the experiment to the models used for the classification purpose. The 3rd section is the evaluation methods where we discussed various evaluation methods that we used to evaluate the performance of the model, also the various techniques we tried to improve the performance of the model. The final Section concludes the results of the project, the suggestions that came up from this study is also mentioned there.

1.2. Previous Studies:

There has been a significant amount of research done on true memory and false memory. Additionally, many studies have been investigated the relation among true and false memory, which turned into the foundation for this project. This section clearly states some of the research paper findings that were examined for this project.

The first experimental analysis of false memories was carried out by Barlett (1932), who is typically given this honour [1]. Reproductive memory refers to accurate, rote output of data from memory, whereas reconstructive memory stresses the active process of filling in missing components while remembering, with errors frequently happening. Barlett (1932) proposed this distinction, which has shown to be enduring. In the list learning paradigm, there is one well-known instance of false memories being created. Underwood (1965) developed a method to analyse false word recognition in lists. He offered the participants a continuous recognition test in which they had to determine if each word was one that had already been presented in a list. Later words had several connections to earlier words that had been examined. Underwood demonstrated that words associated with words that had previously been presented were incorrectly identified.[1]

Patients with a diagnosis of schizophrenia frequently struggle with episodic memory [2,] recall, and recognition. Studies from the past have shown that people with schizophrenia make decisions with high levels of confidence while having less data (Garety et al. 1991; Dudley et al. 1997; Moritz & Woodward, 2004, 2005). Patients may be able to make decisions with a high degree of confidence based solely on contextual cues, such as the towel (lure): "People on the beach normally lie on towels, thus I'm sure I saw plenty." Therefore, it was anticipated that the addition of potent contextual signals would both raise the prevalence of false memories in patients as well as their confidence in comparison to controls. [3].

It is unknown how to distinguish between high-confidence true and false recognition. The current study [4] employed eye movements to measure how well and how confidently people were able to recognise items and themes that were retrieval signals. The gaze pattern may distinguish between True and false recognition with high confidence. In comparison to high-confidence true recognition, high-confidence false recognition was linked to longer fixation times on the item and theme as well as more regression counts on theme [4]. It may be possible to determine if a high-confidence recognition is true or false by monitoring eye movements during the identification.

For the eye-tracking data, a fixation was determined with a combined velocity and acceleration algorithm (Stampe, 1993). Strong memories are processed rapidly and with a high probability, while weak memories are processed more slowly (Wixted et al., 2018; Wixted & Wells, 2017). Even while certain weak memories might be able to reach the subjective threshold for high confidence, they might necessitate prolonged retrieval-related monitoring during item and theme retrieval.

The length of every fixation on the IA (interest area) was added up to determine the fixation duration, which is expressed in milliseconds. There was a stronger correlation between high-confidence false recognition than true recognition and longer fixation length on item and topic as well as more regression counts on theme. The combination of continuous remembrance (specific details connected to a preceding presentation of an item) and familiarity (a sense of familiarity in the absence of contextual data) signals that are returned in response to a retrieval cue determines the memory strength.[4].

1.3. System Requirement

The basis for memory attributions includes a variety of qualitative aspects of the mental experience. For instance, visual, geographical, temporal, or emotional characteristics are frequently considered to be proof that a mental experience accurately represents a memory. These are determined using flexible standards (which qualities are considered and how they are weighted, how much evidence of any given type is needed). As a result, what would be considered a memory in one set of circumstances may not be in another. Most of the time, we would be even going so far as to suggest that most memories are just illusions. This is due to the fact that the memories only bother to retain a small portion of what actually experience and that every time we recall anything, and have the capacity to alter the memory are now accessing.

Accurately recalling any experience, whether it was visual, verbal, or otherwise, is the mark of a true memory. False memory is the recall of an event that didn't happen or a misremembering of one that did. Predicting True and False Memories utilising eye tracking movement from various kinds of machine learning models is one of the main needs for the system. To determine if a memory is genuine or false, we will present the slide, pose questions based on it, and monitor eye movements and reaction times.

The categorization procedure consists of two steps: testing and training. The machine learning model is fed a sizable quantity of data during training in order to create generalised rules. The qualified classifier's final performance is gauged by the classification model. During the testing phase, fresh data is added to mimic a new input class, and the classifiers are guided by the classification model.

The windows 10 operating system was used to develop this project, and the code was written in the Python programming language using libraries to implement the functionality. Team members utilise Python for pre-processing and training machine-learning models. The Python libraries that will be utilised in this project are listed below.

Python Libraries used

Tasks	Libraries
Dataset Reading and Pre-Processing	Pandas
Machine Learning Models for Training and Testing	Scikit-Learn
Data visualization	Matplotlib
For Array-processing	NumPy

Table:1

Software Specification:

This project provides information on the piece of software, or application is known as a software requirements specification (SRS). The terms used to describe the programming languages, codes, and messages needed to communicate with each interface and piece of hardware include (programming interfaces). The software specification step's objectives are to explain the requirements and to get the software application ready for user intervention and validation. The requirement description is used to construct a specification file for a software programme. All the tests that must be performed to complete the software application's validation are described in a specification document for software validation testing.

The software interfaces listed below will be utilised by this project:

SOFTWARE	VERSION
Operating System	Windows 10 or window 11
Python	Python 3.10
Jupyter Notebook	V6 4.11
Python Libraries	Pandas, NumPy, Matplotlib, Scikit-Learn

Table :2

Hardware Specification: -

A hardware compatibility specifies the hardware requirements for the PCs in deployment phase. In general, a hardware specification is built on the results of a sizing study that considers the number of users, the resources needed by each component, and the volume of interactions with each component that require a response. A system specification is a preliminary assessment of the hardware and software capabilities of the system. Every computer user should check to see if the application they wish to install will fit on their system or device and be able to operate on it.

The following hardware interfaces will be used in this project:

HARDWARE	DEVICE DETAILS
System	Laptop/Computer
Eye tracking device	Tobii Pro
Processor	Intel® Core™ i5, i7 or AMD Athon XP 1800,
RAM	Above 8 GB
Hard Disk	40 B Maxtor

Table : 3

Device specification:

A person's presence may be detected by the sensor technology known as eye tracking, which can also track their gaze in real time. The system transforms eye motions into a data stream that includes details like pupil position, gaze vectors for each eye, and gaze points.

Eye tracking Device Specifications	
Device	Tobii Pro Nano
Eye tracking technique	Video-based pupil- and corneal reflection eye tracking with dark and bright pupil illumination modes.
	One camera capture images of both eyes for accurate measurement of eye gaze and eye position in 3D space, as well as pupil diameter.
Sampling frequency	60 Hz
Precision1	0.10° RMS at optimal conditions2
Accuracy1	0.3° at optimal conditions
Binocular eye tracking	Yes
Eye tracker latency	1 frame (17 ms)
Blink recovery time	1 frame (immediate)
Gaze recovery time	250 ms
Data output	Timestamp
	Gaze origin
	Gaze point
	Pupil diameter
Eye image data stream	Not available
TTL input stream	Not available

Table : 4

3. Methodology

3.1. Experiment:

The Experiment conducted is based on the DRM paradigm, 23 participants from different background and age group were voluntarily participated for the experiment. All off them had a normal or corrected-to-normal eyesight with no prior history of neurological or psychiatric diseases. The participants were shown a scene related to their daily life and asked them to remember the objects that they saw in the image and then a series of slides with different object names related to the scene is presented to them. The task given to the participants was to identify the objects that they actually seen in the previous scene.

The purpose of this experiment is to identify the true memory or false memory by using three different emotions-based images. These images are chosen based on three moods of our life.

Here, this beach image shows the positive mood, it would represent the happy emotions.

Fig:1: Beach Scene [11]

Below one represents the negative mood and it triggers the sad emotions.

Fig:2: Accident Scene [11]

The classroom scene represents neutral emotion:

Fig:3: Classroom Scene [11]

Eight object names, that were related to the scene is listed down. In that 8 objects, 4 of them were actually presented in the scene and the other 4 which were related to the scene but not shown in the image. The objects names were written in a slide, with a boundary box and placed in the centre of the screen. These slides were presented to the participants in a shuffled order. During the encoding phase the three visual scenes were presented to the participants for 40 secs each and during the retrieval phase the objects slides were shown with a fixation cross between them, with a time interval of 2 secs. The subjects were then asked to give there response by pressing ‘Y’ and ‘N’ keys in the keyboard, in which ‘Y’ indicates a true response, means they have seen the object in the scene and ‘N’ indicates a negative response. The themes were presented in the following order: Beach, classroom and Accident. All the participants were shown the same images and objects in the same order. Also the calibration of the eye tracking device was done for all he participants before the experiment.

The device used to track the eye movement was tobii pro nano and the software tobii pro lab is used to design and analyse the experiment. The AOI and TOI’s were defined in each slides prior to the test and each experiment is recorded and saved the parameters in the participant name.

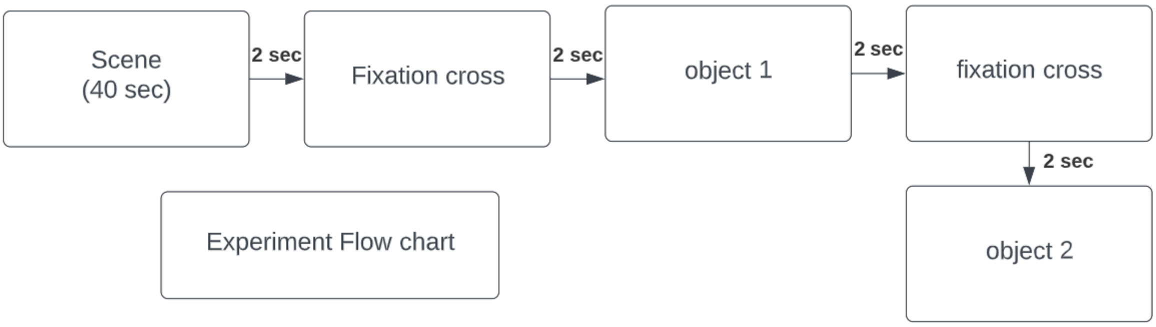

Fig:4: Experiment Flow Chart

3.2. Dataset:

After completing the experiment in tobi pro lab, the data was exported into local drive in excel format. The data exported has 12 columns and 506 rows. Each row represents the response of participants and the parameters corresponding to each object.

Fig:5: Dataset from Tobii lab

TOI stands for time of interest, that is the time in which the features were recorded. Actual response indicates, whether the object is actually present in the scene or not, here the values are Y for objects that were present in the scene and N for objects that were not present in the scene. Similarly, the User Response gives the response of the user, y means he/she seen that object and n for the objects that they think, were not present in the scene. The parameter values like Maximum duration of fixations, No of fixations, Average pupil diameter etc were given in the following columns. In that the total duration of the visit measures the reaction time of participants for each objects.

Now the objective is to classify the datapoints into false and true memory. For that consider the two columns : Actual Response and User Response. If the value for both columns are y or n that means the participants correctly identified the objects, whether they are in the scene or not. And if the value of actual response is N and user response is Y, the participant identified the objects as they were in the scene, but it was not. It can be represented as TP, TN, FP and FN.

Actual Response	User Response	Class category
Y	Y	TP
Y	N	FN
N	Y	FP
N	N	TN

Table:5

The true positive and true negative represents the true memory and false positives are the False memory, So these 3 class categories were retained and the False negatives were eliminated for the sake of our study. In column wise all parameter features were maintained and the features like recordings, TOI, and intervals which will not add any values were removed from the dataset. The final dataset has 463 datapoints in total, with the features: Total Duration of Fixations, Average Duration of Fixations, Maximum duration of fixations, No of fixations, Average pupil diameter, and Total Duration of visit. The class categories TP and TN were converted into True memory and FP class is converted into False memory. The class for True memory is denoted as 1 and for false memory it is 0. The final dataset is shown below.

	Total Duration of Fixations	Average Duration of Fixations	Maximum duration of fixations	No of fixations	Average pupil diameter	Total Duration of visit	class
0	933	467	864	2	2.82689	1100	1
1	3033	3033	3033	1	2.78837	3033	1
2	1266	422	916	3	2.87316	1932	1
3	3684	1228	2725	3	2.96831	5217	1
4	1667	833	1414	2	2.96048	1683	1

Fig:6: Final Dataset

3.3. Feature engineering and EDA:

After uploading the dataset and converting into pandas dataframe, the next step is data pre-processing. This is the data cleaning process that deals with the issues like null values, missing values etc. The details of the data set is summered below.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 463 entries, 0 to 462
Data columns (total 7 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Total Duration of Fixations    463 non-null    int64
 1   Average Duration of Fixations  463 non-null    int64
 2   Maximum duration of fixations  463 non-null    int64
 3   No of fixations                463 non-null    int64
 4   Average pupil diameter         463 non-null    float64
 5   Total Duration of visit        463 non-null    int64
 6   class                          463 non-null    int64
dtypes: float64(1), int64(6)
memory usage: 25.4 KB
```

Fig:7: Dataset Description

As described above the data is very clean, there are no Null or missing values in the features. Also all the features are numerical data, so there is no need to apply any imputation methods. Now we have to check the class distribution among the data points, that is identifying how many data points are in each class.

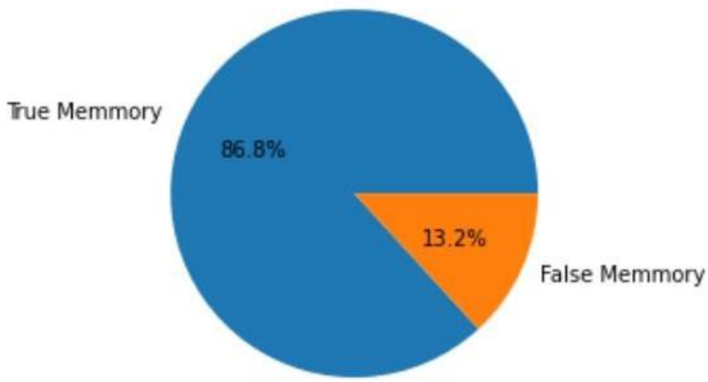

Fig:8: Class Distribution

A pie diagram is plotted for the better understanding of class distribution. From the diagram it is clear that the dataset is highly imbalanced, since the 86.8% of the dataset belongs to a single class. It is necessary to balance the dataset before feeding it into the classification model, otherwise the model will be biased towards the Majority class. Since the datapoints in the minority class is comparatively very less, the under-sampling method is used to balance the dataset. In this method all the data in the minority classes are maintained and the majority class size is reduced. The minority class here is False memory and Majority class is the true memory. After the Under Sampling the dataset will be a perfectly balance one.

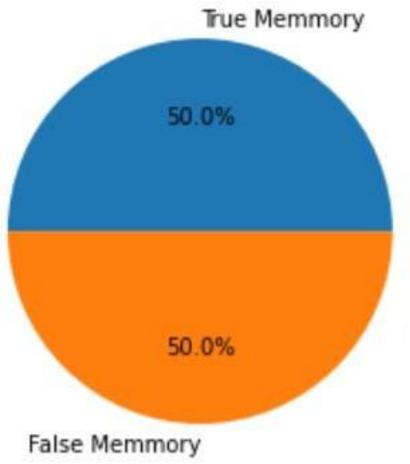

Fig:9: Balanced Class Distribution

In the next step we have to identify the features that are more relevant in predicting the output class and the irrelevant features can be removed from the dataset. Even though the dataset has only 6 features, some features seems insignificant in prediction. In order to find out the important features, the dimensionality reduction method called principal component analysis is applied on the dataset. It will calculate the weightage of each feature in predicting the output class.

	PC1
Total Duration of Fixations	0.544771
Average Duration of Fixations	0.033320
Maximum duration of fixations	0.160495
No of fixations	0.000946
Average pupil diameter	0.000015
Total Duration of visit	0.822408

Fig:10: Feature Weights

Only 3 features has the highest values in PCA, and the features like No of Fixations, Average pupil diameter and Average duration of fixations has a negligible weightage values and removed from the dataset. Finally, the dataset will have 3 features and 2 class labels with 61 data points in each class.

Now we are doing some Exploratory data analysis on the dataset in order to identify how the data are spread out. There are many techniques to do the EDA, here the box plots or whisker plots are used for the analysis. Box plots will show the distribution of numerical data and skewness by plotting the data percentiles. The box plot of each feature against the two class labels are shown below.

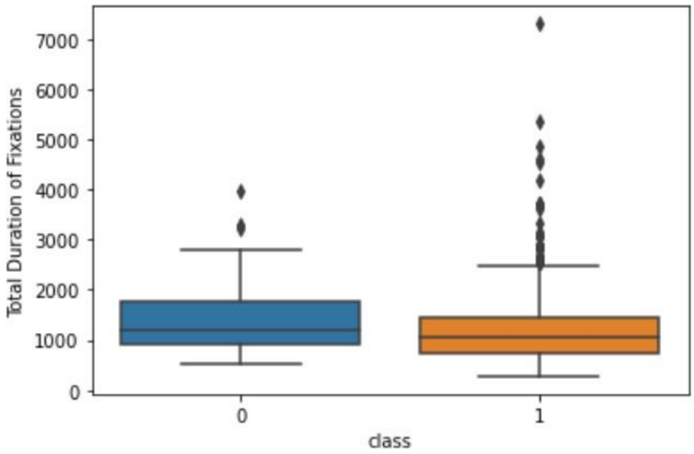

Fig:11: Class Distribution of Features

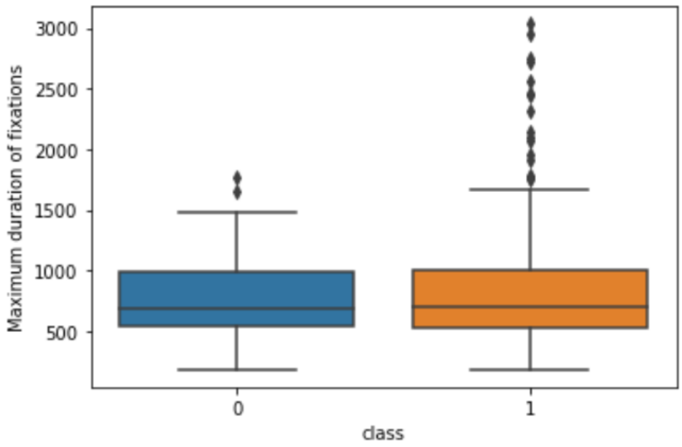

Fig:12: Class Distribution of Features

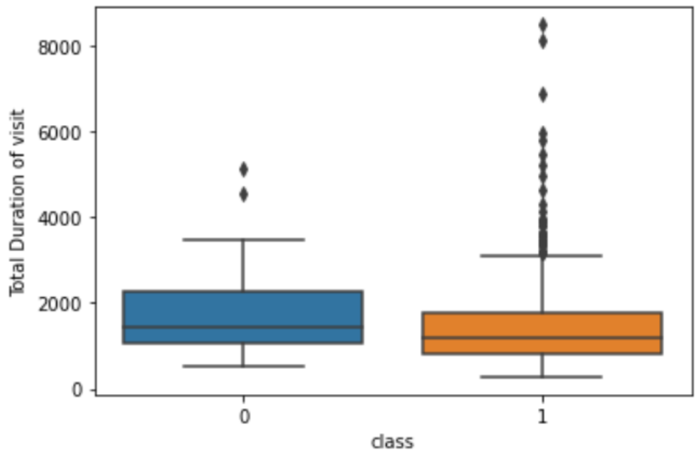

Fig:13: Class Distribution of Features

By analysing the plots, in total duration of visit the value range of the false memory is higher than true memory that implies the participant took more time to react to the false memory objects, that is the reaction time is more for the false memory. Similarly in the case of total duration of fixations the value range is higher for false memory compared to true memory. It is noticeable that the median value for false memory for both the above-mentioned features is a bit higher than true memory. For the feature maximum duration of fixations the 25th percentile and the median values of both the classes are almost similar but there is a slight difference in the 75th percentile. Also, It is noticeable that the true memory class have a lot of outliers or noises in all the three features.

3.4. Classification Model :

Since the class labels are known and there are only two class categories, this is a supervised learning -binary class classification problem. The popular classification algorithms like Logistic Regression, K-Nearest Neighbours, Decision tree and Random Forest were used for the model creation. The dataset is first divided into dependent and independent variables. By applying holdout cross validation the data is split into training and testing data in the ratio 80:20 at a random state of zero. The values of the features lay in different range and a Standard scalar function is applied to standardize it. This function will standardize the feature values by removing the mean and scaling to unit variance. The scaling is done for both training and testing data. The classification models were trained on the training data and the performance is evaluated later, by using the test dataset.

Logistic Regression: In this project two types of classification algorithms, both the linear and no linear were used. The linear algorithm used here is Logistic regression which is very efficient in predicting the binary class classification problems. It uses a sigmoid function to predict the probability of the class based on a set of given features, since it will give the prediction as a probability, the output will be in the range zero to one and the direction of association can also identified from LR. Since we have a low dimensional data, the chance for overfitting is comparatively low.

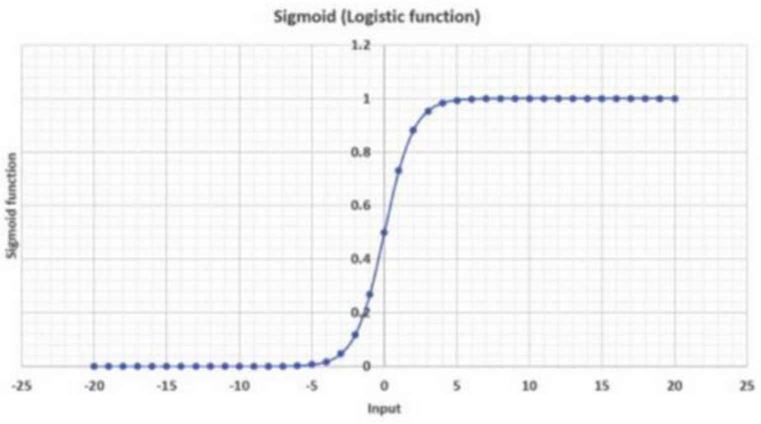

Fig:14: Sigmoid Function In LR[8]

KNN: The KNN is a non parametric classification algorithm which will work based on the assumption that the datapoints with similar properties will grouped together. The predictions are made based on majority vote and a majority greater than 50% normally applied for binary classifications. In order to find out the closest data points, the distance between the data points were calculated, there are different distance measurements that we can use, here we are using the default distance metric minkowski and a default neighbour size 5 for the classifier.

Decision Tree: The other non linear model used here is Decision trees , which is based on the intuition that the classifier will make yes/no questions based on the dataset features and spilt the dataset continuously until all the datapoints are classified into either one of the classes. The classification process starts from the roots of the tree, then values of the root attributes is compared with the attributes of the record and based on the comparison, a branch is created and will move to the next node. The parameters of the decision tree classifier was set to default values.

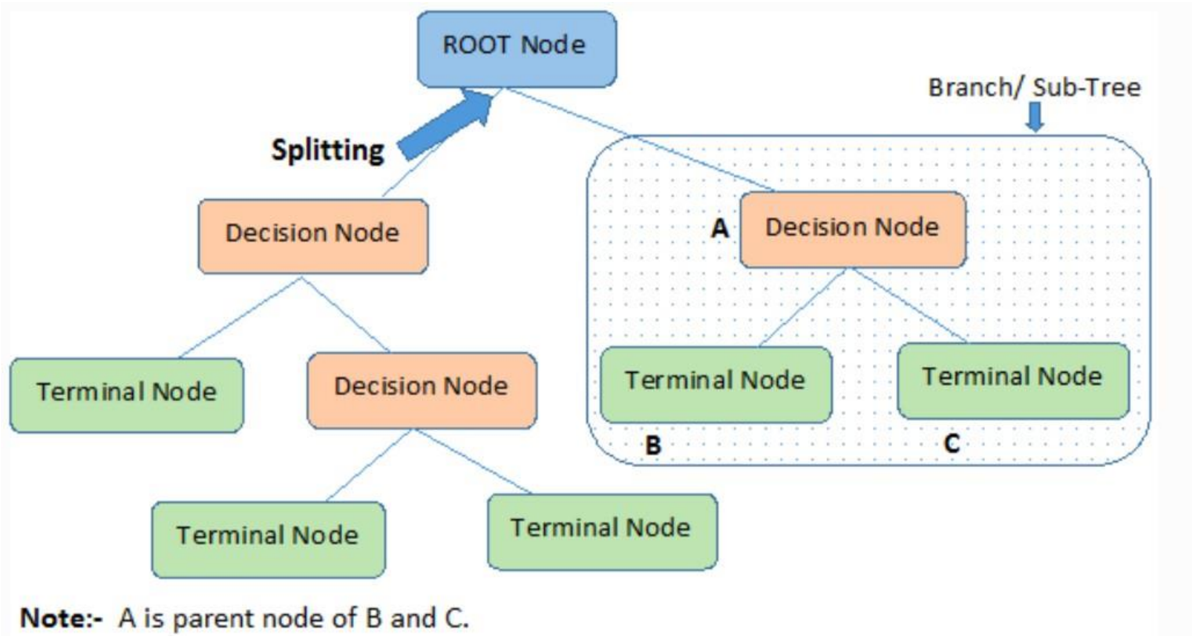

Fig:15: Decision Tree[9]

RandomForest: An ensemble model Randomforest, which is constructed based on decision trees was also used for the classification task. For ensemble the multiple decision tree models the Bootstrap aggregating or bagging method is used in randomforest. In this method the prediction are made with multiple DT models and the final prediction is made by taking the mean of all the individual predictions or by majority vote. The number of trees used here is 100 by default and the gini impurity is applied to measure the quality of the split.

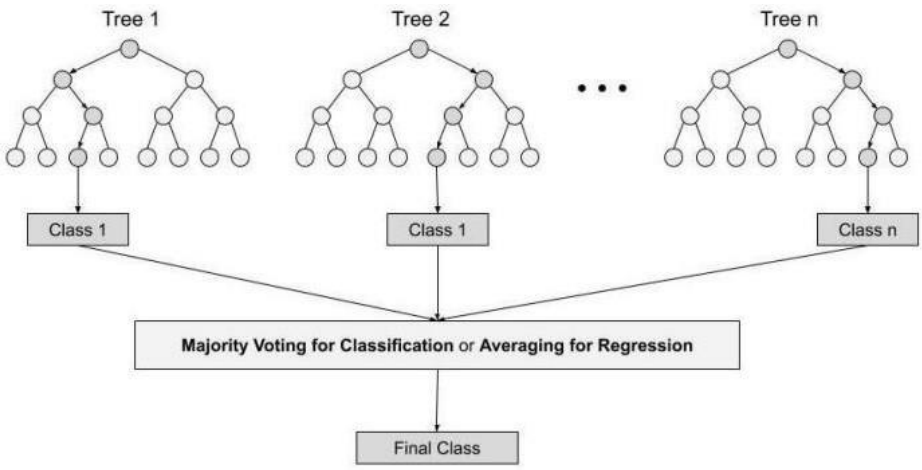

Fig:16: Random Forest Classifier[10]

4.Evaluation:

After training the model in training dataset, the model performance is evaluated using the test dataset. The classification model performances were evaluated using the metrics like Accuracy score, precision, and recall. Accuracy is defined as the ratio of total no of accurate predictions to the total predictions made. Precision is measured as the ratio of true positive predictions to the total no of positive values. The recall value indicates the ability of the model to identify an observation that belongs to the positive class randomly.

$$Accuracy = \frac{No\ of\ correctly\ classified\ samples}{Total\ no\ of\ classifications}.$$

$$Precision = \frac{TP}{TP\ +\ FP}.$$

$$Recall = \frac{TP}{TP\ +\ FN}.$$

Individual accuracy scores of all the classifiers are calculated and the classification report of the classifier which gave the maximum accuracy value is created. Then a confusion matrix is plotted in order to represent the data graphically. The accuracy score values of the classifiers are shown below:

```
Logistic Regression  :  : 40.00%
K-Nearest Neighbors  :  : 48.00%
      Decision Tree  :  : 60.00%
      Random Forest  :  : 68.00%
```

Fig:17: Accuracy score values of Classifiers

The randomforest classifier has the maximum accuracy value of 68% followed by decision tree with an accuracy of 60%.Logistic regression has the least value of 40%.Now the classification report is made for the Randomforest classifier in order to calculate the precision and recall .

```
Classification Report for RandomForestClassifier :

              precision    recall  f1-score   support

           0       0.64      0.64      0.64        11
           1       0.71      0.71      0.71        14

    accuracy                           0.68        25
   macro avg       0.68      0.68      0.68        25
weighted avg       0.68      0.68      0.68        25
```

Fig:18: Classification report of Random forest classifier

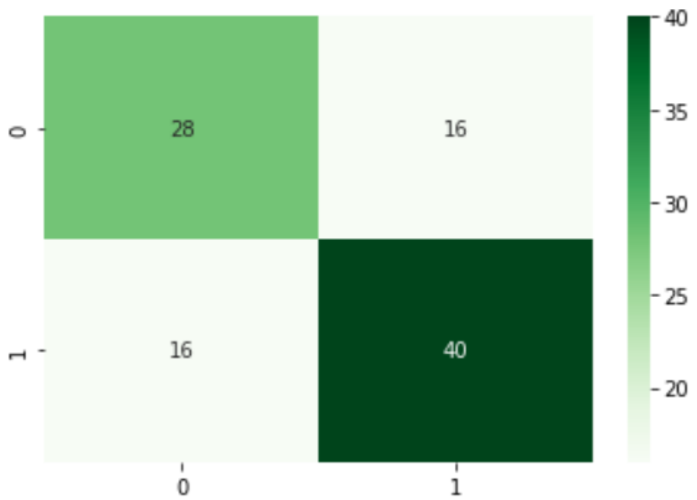

Fig:19: Confusion Metrics

The classifier well predicted the true positive values but the prediction accuracy of the false memory is comparatively low. Both the false positives and true negatives has the same values, which is high. Which indicates the classifier predicted a reasonable number of test data wrongly. In the next step we are attempting to increase the accuracy of the classifier and there by optimising the performance.

4.1. Weighted Features:

After the feature engineering, the dataset has 3 features which are Total Duration of Fixations, Maximum duration of fixations and Total Duration of visit. The weightages of each of these features was calculated during the Principle component analysis. Now these weights are multiplied with the values of the corresponding features and then these feature were summed together to create a new feature called Weighted values.

	weighted value
0	4568.032523
1	1673.573863
2	1790.517198
3	6482.951150
4	2765.846921
...	...

Fig:20: Weighted feature values

This single weighted feature is used to train and test the model. Its should note that the there were no changes made to the dependent variable. Now we repeat the steps again and the accuracy sore is calculated for each classifiers with the new feature:

```
Logistic Regression  :  : 40.00%
K-Nearest Neighbors  :  : 48.00%
      Decision Tree  :  : 56.00%
      Random Forest  :  : 56.00%
```

Fig:20: Accuracy score values with weighted features.

There is no improvement in the model performance , a box plot is plotted to show the class distribution against the weighted features.

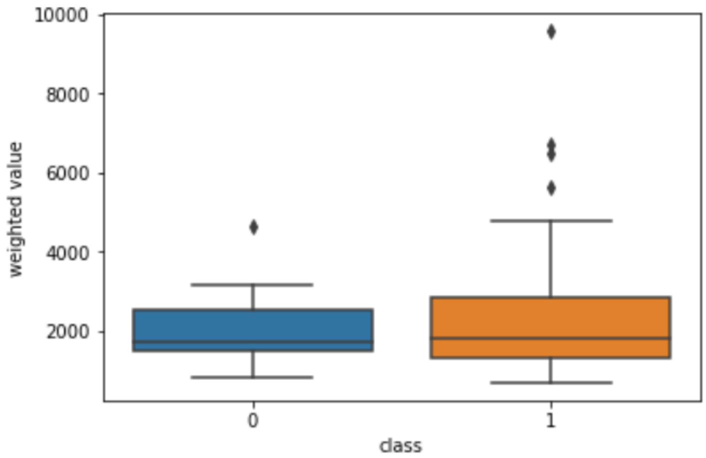

Fig:21: Class distribution of weighted features

By analysing the graph, the values of the false memory class is completely overlapping with the values of the true memory class. so the classifier failed to distinguish the two output classes with this weighted feature. In the next step we are trying to check whether there is a threshold value for the weighted feature that will separate the two classes. A simple classifier is created for the purpose.

4.2. Threshold Value:

An attempt is made to find whether there is a threshold value for the weighted feature that can well separate the classes. A simple classifier was created, which will work such a way as each time it will pick one value from the weighted feature as threshold and predictions were made for the values less than the threshold as zero and above as one. Finally the accuracy against each threshold values were calculated and displayed in descending order using pandas data frame .

	weighted value	accuracy
49	3662.779720	0.532787
61	3782.134477	0.524590
55	3411.785538	0.524590
35	3374.836028	0.524590
23	3616.071669	0.524590
...	...	...

Fig:22: Accuracy values for individual thresholds

The value 3662.77 gave the maximum accuracy value, but while checking the feature column, only 10 % of data came below this particular value. Therefor we cannot set this as a threshold for separating the two classes.

5.Discussions and Conclusions:

As we discussed earlier, the experiment used here is similar to the DRM paradigm and the eye movements were recorded with the help of a tobii pro eye tracker. The recordings were precise, and it was easy to convert the data into any kind of format as per our convenience. Even though there were 3 different emotional scenes used for the experiment, the parameters recorded didn’t show much variation with the emotions. Maybe there is a chance that the scenes may not trigger the desired emotions in the participants. As per the previous experiments the scenes were shown to the user for 40 secs, the images used in those experiments had more object compared to the scenes used for this study. This may helped the user to remember the objects much easier, so the suggestion is to adjust the time for each scenes based on the complexity of objects present in the image. Most of the participants were able to give the correct response, resulting in more data points in the true memory class compared to the false memory. Even though the participants are from different backgrounds and age groups, we didn’t take that in to consideration while collecting the dataset, Also it is observed in the experiment that there was not much corelation between the parameters and the background of the people. The participants gave mixed response irrespective of the age and their ethnicity. A suggestion from the experiment side is to use a more complex images and adject the scene timings based on that.

Even though we got a decent accuracy of 68%, the features used in the dataset were not able to completely separate the two classes. The precision and recall values are just fair enough and still there are a lot of points miss classified as True and false memory. Different feature engineering and EDA methods used in the study to extract the required information from the data, but they were not successful in improving the model performance. The imbalance in the dataset maybe a reason for this since there are only a few data points in the false memory class. So the suggestion from the study is to add more data points in the false memory class and create a balanced dataset. Also it is very much encouraged to include other features such as saccades time, movements etc into the dataset, thereby increasing the dimensionality of the features. Another suggestion is to add more class categories in the experiment such as high confidence false memory, low confidence true memory etc .This may also help to classify the data points more efficiently.

References:

1) Roediger, Henry L., and Kathleen B. McDermott. "Creating false memories: Remembering words not presented in lists." Journal of experimental psychology: Learning, Memory, and Cognition.
2) Tulving, E. (2001). Episodic memory and common sense: how far apart ? Philosophical Transactions of the Royal Society of London, Series B Biological Sciences 356, 1505–1515.
3) MORITZ, S., WOODWARD, T., & RODRIGUEZ-RAECKE, R. (2006). Patients with schizophrenia do not produce more false memories than controls but are more confident in them. Psychological Medicine.
4) Dang, Xixi, et al. "Distinguishing between high-confidence true and false memories: evidence from eye movements." Australian Journal of Psychology 73.2 (2021): 243-253.
5) https://www.intechopen.com/chapters/64920 9.
6) https://nobaproject.com/modules/eyewitness-testimony-and-memory-biases 10.
7) https://www.tobiipro.com
8) https://www.sciencedirect.com/topics/computer-science/logistic-regression
9) https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html
10) https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest
11) https://www.shutterstock.com/