MEF UNIVERSITY

PREDICTING CUSTOMER PERCEPTION ON BRANDS USING FUNCTIONAL NEAR-INFRARED SPECTROSCOPY MEASUREMENTS

Capstone Project

Emre Kemerci

İSTANBUL, 2019

MEF UNIVERSITY

PREDICTING CUSTOMER PERCEPTION ON BRANDS USING FUNCTIONAL NEAR-INFRARED SPECTROSCOPY MEASUREMENTS

Capstone Project

Emre Kemerci

Advisor: Asst. Prof. Utku Koç

İSTANBUL, 2019

MEF UNIVERSITY

Name of the project: Predicting Customer Perception on Brands Using Functional Near-Infrared Spectroscopy Measurements
Name/Last Name of the Student: Emre Kemerci
Date of Thesis Defense: 09/09/2019

I hereby state that the graduation project prepared by Emre Kemerci has been completed under my supervision. I accept this work as a "Graduation Project".

09/09/2019
Asst. Prof. Utku Koç

I hereby state that I have examined this graduation project by Emre Kemerci which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

09/09/2019

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of Emre Kemerci and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1. Asst. Prof. Utku Koç	……………………
2. Prof. Dr. Özgür Özlük	……………………

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Emre Kemerci 09/09/2019 Signature

EXECUTIVE SUMMARY

PREDICTING CUSTOMER PERCEPTION ON BRANDS USING FUNCTIONAL NEAR-INFRARED SPECTROSCOPY MEASUREMENTS

Emre Kemerci

Advisor: Asst. Prof. Utku Koç

SEPTEMBER, 2019, 36 pages

Customer perception on the brands have importance to give strategic decisions by marketing professionals. In classical ways, customer perception on brands are researched through conducting field surveys. Similarly, neuromarketing discipline have studies on customer behaviors, their perceptions, communication techniques etc. under the frame of decision-making process of human. In neuromarketing, functional near-infrared spectroscopy (fNIRS) is a technology used to measure oxy and deoxy hemoglobin concentration in the tissues in order to enable to analyze hemodynamic responses of the brain activities. In this study, a group of participants' activations of prefrontal cortex so the hemodynamic responses that were collected against a set of stimuli, which is a brand logo and adjective associated with the brand is used as dataset. Measured hemodynamic response metrics are oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR), total hemoglobin (HbT) and Oxygenation (Oxy) and the dataset includes 168 participants' measurements for 30 stimuli. In addition, the information regarding the responses of the participants and common perception of stimuli (field study results for same stimuli) are also exists in dataset. The aim of the project is to predict through machine learning algorithms whether relation between brand and the relevant adjective is Positive, Negative or Neutral using these feature set. As methodology of this study, fNIRS measurements in the data is cleaned and Null values are handled, measurements are consolidated per participant and stimuli with two different method as feature creation and classification algorithms are used as supervised learning to predict brand perception. In conclusion, performance of support vector classifier and XGBoosting algorithms are become very low, slightly over 50% accuracy despite the optimization with different classifier parameters. Further studies are addressed as performing feature engineering studies with different options.

Key Words: fNIRS, Brand Perception, Neuromarketing, Prediction, Machine Learning.

ÖZET

FONKSİYONEL YAKIN KIZILÖTESİ SPEKTROSKOPİ ÖLÇÜMLERİ KULLANILARAK TÜKETİCİLERİN MARKA ALGILARININ TAHMİNLENMESİ

Emre Kemerci

Tez Danışmanı: Yrd. Doç. Utku Koç

EYLÜL, 2019, 36 sayfa

Markalar üzerindeki tüketici algısı, pazarlama profesyonellerinin stratejik kararlar vermesi açısından önem kazanmaktadır. Klasik yöntemlerde, markalar üzerindeki tüketici algısı saha araştırmaları ile incelenmektedir. Benzer şekilde nöropazarlama disiplini de insanların karar alma süreci çerçevesinde tüketici davranışları, algıları, iletişim teknikleri vb. konularda çalışma yapmaktadır. Nöropazarlama alanında beyin aktivitelerinin hemodinamik karşılıklarını analiz edebilmek üzere dokulardaki oksijenleşen ve deoksijenleşen hemoglobin konsantrasyonunu ölçmek üzere fonksiyonel yakın kızılötesi spektroskopi (fNIRS) teknolojisi kullanılmaktadır. Bu projede, bir grup katılımcının, marka logosu ve bu markaya ithaf edilmiş sıfattan oluşan bir uyaran setine dayalı prefrontal korteks aktivasyonlarını, dolayısıyla hemodinamik sonuçları veri seti olarak kullanılmıştır. Veri seti 168 katılımcının maruz kaldıkları 30 uyaran neticesinde oksijenleşen hemoglobin (HbO), deoksijenleşen hemoglobin (HbR), toplam hemoglobin (HbT) ve oksijenlenme (Oxy) metriklerini içeren hemodinamik tepkileridir. Ek olarak katılımcıların verdiği cevaplara ilişkin bilgiler ve uyaran hakkında genel kabul görmüş algı da (aynı uyaranlar için saha çalışması sonuçları) veri setinde bulunmaktadır. Projenin amacı, makine öğrenmesi algoritmaları ile marka ve marka ile ilişkilendirilmiş sıfat arasındaki ilişkiyi Pozitif, Negatif ya da Nötr olarak sınıflandırarak tahminlenmesidir. Çalışma metodolojisi olarak veri setindeki fNIRS ölçümleri temizlenmiş, boş değerler ele alınmış, nitelik oluşturmak açısından katılımcı ve uyaran bazında ölçümler 2 farklı metot ile konsolide edilmiş ve marka algısını tahminlemek üzere güdümlü öğrenme ile sınıflandırma algoritmaları kullanılmıştır. Çalışma sonucu olarak, destek vektör sınıflandırıcı ve XGBoosting algoritmalarının performansı yapılan model optimizasyon çalışmalarına rağmen %50'nin çok az üzerinde doğruluk ile çok düşük belirmiştir. İlave çalışmalar olarak, farklı nitelik geliştirme işlemlerinin uygulanması gerekliliği adreslenmiştir.

Anahtar Kelimeler: fNIRS, Marka Algısı, Nöropazarlama, Tahminleme, Makine Öğrenmesi

TABLE OF CONTENTS

1. INTRODUCTION

The human decision-making process does not fully rely on rational calculations and awareness on all pros and cons within the alternative decisions. This also sounds for purchasing behavior of human. From the theory of Adam Smith, claiming that the consumers are rational for their buying decisions, today cognitive sciences try to deal with perceptual effects on human decisions in buying decisions. Considering the rational and cognitive components of decision making in the buying process, on the basis of neuroscience, economics and human psychology, neuromarketing rises as an essential field of marketing area in order to understand the human decision-making process. (Sebastian, 2014)

Neuroscience deals with how the nervous system functioning and since it involves several disciplines also have interfaces with areas like biology, physiology, marketing etc. The use of neuroscience varies multiple high-tech applications which are all types of measurement of brain activities like EEG (electroencephalogram), MR (magnetic resonance), PET (positron emission tomography). These techniques require high-cost large scanners and they need that participants being positioned in these devices; in other words, they are not suitable for mobilization. (Kopton & Kenning, 2014)

Against the limitations of these techniques, fNIRS (Functional near-infrared spectroscopy) is also used to measure brain activities with its compatibility in field studies. fNIRS is an optical brain imaging technique without any harm to body and used to measure cerebral blood flow (CBF) as well as the hemodynamic response in a local brain area during neural activity. Although, functional magnetic resonance imaging (fMRI) is used as a standard in neuroimaging, fNIRS is also reliable and valid measurement for cortical activations. (Kopton & Kenning, 2014)

In fNIRS technology, near-infrared light (with a wavelength spectrum of circa 650–950 nm) that is reaching to tissue is used as source and near-infrared detectors to measure the reflected infrared light. In tissue, oxy-(O_2Hb) and deoxy hemoglobin (HHb) absorb the infrared and since the absorption are not isodense for them, changes in oxy- and deoxy-hemoglobin concentration can be monitored and assessing the neural activity is possible. (Kopton & Kenning, 2014) Via fNIRS, cortical hemodynamic responses are obtained applying sources/detectors over the scalp. Infrared can affect and detectors so measure approximately 1 cm under the scalp and 10 measurements can be taken per second typically.

Illustration of the development of a typical cortical hemodynamic response with a stimulus is given in Figure 1 below: (Quaresima & Ferrari, 2016)

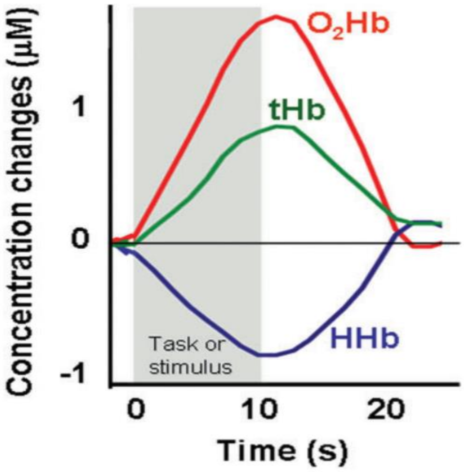

Figure 1: Typical development of hemodynamic response

The home for decision making, judgment, reasoning, emotional regulation, social–relational abilities, and abstract thinking in the brain is the prefrontal cortex, the front part of the frontal lobe. Since the buying decision is managed by the prefrontal cortex, measurements with fNIRS are taken from the frontal part of the head.

Within this perspective; in this project customer perception of the brands against the attributes associated with these brands are predicted via machine learning models through analysis of hemodynamic responses. In the experiment, 30 brands' logo and adjectives like small, laughy, salty, sportive etc. associated with them are used as stimuli set that are exposed to experiment participants. Participants' hemodynamic responses, which is measured with fNIRS device, explicit answers of participants given to relation between brand – adjective match and common public perception about the matching obtained from field research are used in this project.

2. PROJECT DEFINITION

Within the scope of this project, the data is collected via fNIRS (Functional near-infrared spectroscopy) device from a group of participants for a set of stimuli. For this participant group, in an isolated laboratory environment, activation of prefrontal cortex so the hemodynamic responses were collected against a set of stimuli, which is a brand logo and adjective associated with the brand. During the test, participants were asked to answer the question of whether he/she is agreed that each adjective is accord with the brand shown in the screen. The answers are collected as “Yes” (positive), “No” (negative) and “Don’t know” (neutral). For example, with the logo of “Prada” (brand), the adjective “Cheap” are shown on screen and it is expected from participant to select the choice of either “Agree”, “Do Not Agree” or “Have no idea”.

Other than the responses given by participants, we have information obtained from field research which was conducted with same stimuli set and similarly the relation between brand and corresponding adjective is marked as Positive, Negative or Neutral. The aim of the field research is to derive commonly held perceptions of this brand-adjective matches.

The mentioned laboratory studies were performed with 172 participants and 30 stimuli set is used for each participant. The outputs of fNIRS device measurements per respondent and stimuli are stored in txt files while the data related with participants’ answers are stored per respondent in another set of txt files.

The aim of this project is to analyze the data representing the hemodynamic responses of the participants collected via fNIRS system and predicting the brand perception as either positive, negative or neutral using the machine learning algorithms.

The project plan consists of literature review in section 3, brief explanation about the raw data with explanatory data analysis in section 4. The Methodology section is dedicated to data preparation and cleaning studies, feature selection and creation techniques used, applied machine learning algorithms and evaluation metrics. Lastly conclusion and further studies are addressed in section 6.

3. LITERATURE REVIEW

In cognitive neuroscience area, Hosseini et al., 2011 studied on the application of the support vector machine (SVM) as a classification method to predict the human behavior as like or dislike with fNIRS measurements on anterior frontal cortex. The study shows that the measurements taken from the anterior medial frontal region is enabling the classification of human behavior whether the stimuli is liked or disliked. (Hosseini, et al., 2011)

In a different study to examine consumer's responses (via fNIRS) to different merchandising communication strategies (like wall decoration, product displays, etc.) at the point-of-sale (PoS), it is considered that fNIRS is a valid method to measure hemodynamic responses of the prefrontal cortex (PFC) of shoppers, so as the buying decisions. Secondly, the results of the study revealed that orbitofrontal cortex and dorsolateral prefrontal cortex (parts of prefrontal cortex) are the regions handling the merchandising communication strategies. (Krampe, Strelow, Haas, & Kenning, 2018) In addition, the study of Liu X, Kim C-S aiming to classify four store type as negative and positive using hemodynamic responses, worked with a group of statistics of signals like mean, variance, t-value, slope etc. with linear discriminant analysis. Through using t-values derived from data, they reached the conclusion that average accuracy for positive samples are 50.68% where it is 51.07% for negative ones. Also, their study shows that accuracy of women participants for womenswear store reaches to 73%. (Liu, Kim, & Hong, 2018)

Moreover, Luu and Chau were studied with 9 participants and asked to prefer one drink from two choices and measured the hemodynamic responses with FNIRS. They used the mean of measurements and reached an average of 80% accuracy to predict which drink was preferred with linear discriminant analysis. (Luu & Chau, 2008)

Regarding participants' assessment on brand logos, through experiment conducted with Functional Magnetic Resonance Imaging, the frontal orbital cortex and the frontal medial cortex activation is shown both without pre-instructing (implicit) and explicitly. (Santos , Seixas, Brandão, & Moutinho , 2012) So, we know that frontal cortex is the region on brain to work brand perceptions.

Besides, regarding purchasing behavior, analysis of neural activations in the prefrontal cortex with a purchasing decision as "buy" or "do not buy/pass" through FNIRS concluded that neural activities are able to explain buy/pass decisions with 71% accuracy and in case participants behavior on use of their budgets (budget sensitive and budget insensitive), accuracy reaches to 85%. Eventually, the study revealed that fNIRS methodology is able to model predictive model of consumer purchasing behavior. (Çakir, Çakar, Girişken, & Yurdakul, 2018)

4. ABOUT THE DATA

The fNIRS device used to collect neural activation data for my project has 16 optodes (detectors) and 4 light sources to measure oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR), total hemoglobin (HbT; also shows cerebral blood volume) and Oxygenation (Oxy) levels of the tissues under the optodes.

Via using fNIRS device, participants are shown a stimulus for 4 seconds - the period for a mental evaluation, then they asked to select their answers for another 4 seconds and finally, 8 seconds followed for fixation as figured below. The flow of an stimuli is figured below:

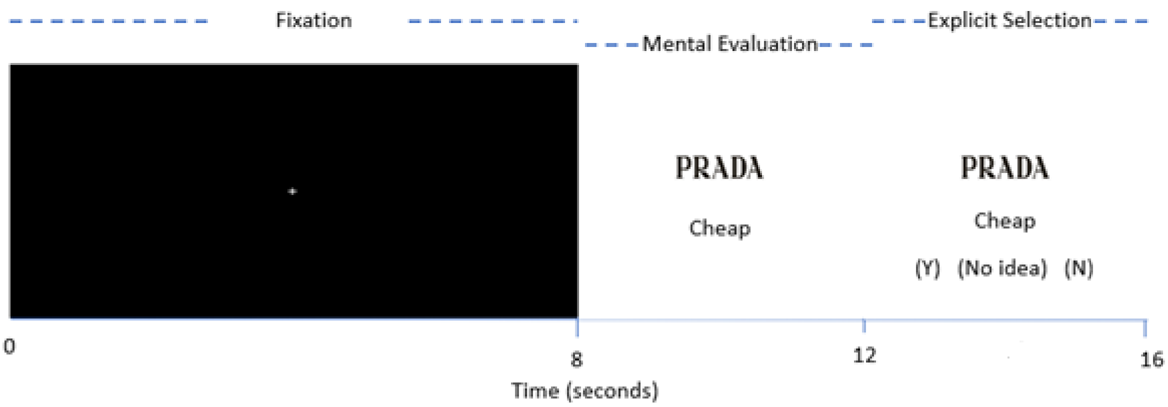

Figure 2: Time capture for a trial of stimuli

Since the fNIRS device works in 2Hz and the mental evaluation and answering phases lasts for 8 seconds, 16 measurements are taken for each stimulus. For each stimulus shown to a participant, the measurements taken from 16 optodes are stored in txt files separately for metrics HbO, HbR, HbT, Oxy. So, for 168 respondents, 30 stimuli for each respondent and 4 metrics for each stimulus, there are 20,160 txt files. containing 16x16 matrix of measurements.

Through reading and appending all txt files mentioned above, a dataframe is created with 17 columns and 325,024 rows. The name of the columns and explanations of them are as follows:

Optode 1 to Optode 16	Measurement values taken from each 16 optodes
Filename	The name of the file of which relevant measurements are read. Path of the filename represented as Participant#_Stimuli#.Metric.Block.txt

Table 1: Explanation of columns of fNIRS device raw data

Since the filename column is a combination of three information, the column is split to three columns as Metric, Participant and Stimuli Number. Then the column “Metric” is converted to features so each optodes’ measurement values were multiplied with four metrics. I also added the columns “measurement”, “millisecond” showing the order of measurement and its equivalent in terms of millisecond for each stimulus. After these preparation steps, fNIRS data has 81,256 observations and 69 columns. The columns of final fNIRS data and descriptions of columns are as follows:

```
fnirs.columns
```

```
Index(['Participant', 'StimuliNumber', 'Hbo_Op1', 'Hbo_Op2', 'Hbo_Op3',
       'Hbo_Op4', 'Hbo_Op5', 'Hbo_Op6', 'Hbo_Op7', 'Hbo_Op8', 'Hbo_Op9',
       'Hbo_Op10', 'Hbo_Op11', 'Hbo_Op12', 'Hbo_Op13', 'Hbo_Op14', 'Hbo_Op15',
       'Hbo_Op16', 'Hbr_Op1', 'Hbr_Op2', 'Hbr_Op3', 'Hbr_Op4', 'Hbr_Op5',
       'Hbr_Op6', 'Hbr_Op7', 'Hbr_Op8', 'Hbr_Op9', 'Hbr_Op10', 'Hbr_Op11',
       'Hbr_Op12', 'Hbr_Op13', 'Hbr_Op14', 'Hbr_Op15', 'Hbr_Op16', 'Hbt_Op1',
       'Hbt_Op2', 'Hbt_Op3', 'Hbt_Op4', 'Hbt_Op5', 'Hbt_Op6', 'Hbt_Op7',
       'Hbt_Op8', 'Hbt_Op9', 'Hbt_Op10', 'Hbt_Op11', 'Hbt_Op12', 'Hbt_Op13',
       'Hbt_Op14', 'Hbt_Op15', 'Hbt_Op16', 'Oxy_Op1', 'Oxy_Op2', 'Oxy_Op3',
       'Oxy_Op4', 'Oxy_Op5', 'Oxy_Op6', 'Oxy_Op7', 'Oxy_Op8', 'Oxy_Op9',
       'Oxy_Op10', 'Oxy_Op11', 'Oxy_Op12', 'Oxy_Op13', 'Oxy_Op14', 'Oxy_Op15',
       'Oxy_Op16', 'Participant_Stimuli', 'measurement', 'milisecond'],
      dtype='object')
```

Figure 3: Columns of fNIRS data

Participant	The participant number/ID
Stimuli Number	The stimuli number/ID
Hbo_Op1 to Oxy_Op16	Measurement values of each metric for each optode from 1 to 16
Participant_Stimuli	The code created artificially by combining participant number and stimuli number
measurement	The measurement number created artificially. Shows the order of measurement for each stimulus.
millisecond	The data derived from measurement column. Since 2 measurement taken per second taken, it shows the millisecond that the measurement taken.

Table 2: Explanation of columns of fNIRS dataset after preparation

As the preprocessing activity of fNIRS data, these measurements are cleaned from motion artifacts, a type of noise arise from respondents muscular, respiratory and other movements. In addition, the measurement values for each stimulus are normalized with the ending values of the preceding one, in other words first measurement of each stimuli shows the change from the ending measurement of previous stimuli.

In addition to hemodynamic measurements explained above, the information regarding the responses of the participants and common perception of stimuli as Positive, Negative and Neutral are also recorded in 171 txt files for each respondent. This dataset is named as “MarkaSifat” data.

After reading and appending these files, the raw data is filtered for the information we need, and the participant number is split from filename. Since the txt files are organized as row by row, the variables are also converted to columns.

Before conversion

	variable	value	Participant2
0	Procedure	EH	1
1	MARKAEH.RT	514	1
2	MARKAEH.RESP	{RIGHTARROW}	1
3	MARKAEH.CRESP	{LEFTARROW}	1
4	MARKA	R03	1
5	Procedure	HE	1
6	MARKAHE.RT	511	1
7	MARKAHE.RESP	{DOWNARROW}	1
8	MARKAHE.CRESP	{RIGHTARROW}	1
9	MARKA	R05	1
10	Procedure	EH	1

After conversion

	Procedure	Participant	MARKA	ResponseTime	Response
0	EH	1	R03	514	{RIGHTARROW}
1	HE	1	R05	511	{DOWNARROW}
2	EH	1	P05	557	{LEFTARROW}
3	HE	1	P07	1082	{RIGHTARROW}
4	EH	1	N03	1368	{RIGHTARROW}

I added the columns “Stimuli Number”, “Response Code”, “Type” and “Type Code” columns to the dataframe. After these preparation steps, the shape of the data is 5130 rows x 9 columns. The names of the columns and their explanations are as follows:

Procedure	The order of choices shown in the screen. EH means the choice of Yes is in left side, HE means the choice of Yes is in right side In the experiment, order of choices changed randomly for each stimulus to prevent lateralization bias - the tendency for some neural functions or cognitive processes to be specialized to one side of the brain or the other.
Participant	The participant number/ID
Marka	The code representing the brand-adjective set / ID
Response Time	The time (in milliseconds) participant responded to the question
Response	The button participant clicked on keyboard to answer the question. It is either Right Arrow, Down Arrow or Left Arrow. Down Arrow corresponds to Neutral Right Arrow corresponds to No in case the procedure is EH otherwise Yes Left Arrow corresponds to No in case the procedure is HE otherwise Yes
Stimuli Number	The stimuli number/ID

Response Code	The code created artificially to convert the column “Response” to numerical values. 1 represents Negative (No), 2 represents Neutral (Don’t Know) and 3 represents Positive (Yes)
Type	The label derived from the column “Marka”. It represents the relation between Brand and Adjective learned from field study – common perception. N represents Negative, R represents Neutral and P represents Positive.
Type Code	The label created artificially from the column “Type” to convert the categories in this column to numerical values. 1 represents Negative (No), 2 represents Neutral (Don’t Know) and 3 represents Positive (Yes) – same coding with Response Code

Table 3: Explanation of columns of” MarkaSifat” dataset

Lastly, I merged second dataframe on the first one using the “Participant” and “Stimuli Number” keys and get the main dataframe of 81256 rows × 76 columns.

Explanatory Data Analysis

The dataset I have includes fNIRS measurements of 168 unique participants for 30 unique stimuli creating 5040 unique participant-stimuli matches as shown in Figure 4.

```
df1["Participant"].nunique()
```

```
168
```

```
df1["StimuliNumber"].nunique()
```

```
30
```

```
df1["Participant_Stimuli"].nunique()
```

```
5040
```

Figure 4: Number of participants & stimuli and their matches

When the number of measurements taken from fNIRS device per participant-stimuli checked, it is seen that for 617 matches 17 measurements are taken in 8 seconds (Figure 5). It’s explaining why number of rows in fNIRS measurement dataset are more than expected.

measurement	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Participant_Stimuli	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5040	5039	617

Figure 5: Number of measurements per participant-stimuli match

In the dataset, there are 3 categories of Type Code and for each one there are 1670 participant-stimuli match, which means every participant is shown 10 positive labeled stimuli, 10 negative labeled stimuli and 10 neutral labeled stimuli since number of participants is 167 for each label. However, for the Response Code, there are 2006 matches selected as Negative, 2182 as Positive and 600 as Neutral. So, through comparing the number of matches for Type Code and Response Code as in Figure 6, 7 and 8, it is seen that consistency of labels between common perception and participant responses are high positive and negative labels while it is so low for neutral label.

TypeCode	Participant	StimuliNumber	Participant_Stimuli
1.0	167	30	1670
2.0	167	30	1670
3.0	167	30	1670

Figure 6: Type Code vs Participant_Stimuli

ResponseCode	Participant	StimuliNumber	Participant_Stimuli
1.0	160	30	2006
2.0	126	30	600
3.0	161	30	2182

Figure 7: Response Code vs Participant_Stimuli

TypeCode	ResponseCode	Participant_Stimuli
1.0	1.0	1358
	2.0	136
	3.0	101
2.0	1.0	563
	2.0	382
	3.0	644
3.0	1.0	85
	2.0	82
	3.0	1437

Figure 8: Type Code vs Response Code

Regarding the measurement values obtained from each optode, how the values are distributed is reviewed. There is concentration around zero and it is seen that distributions have similar shapes for each metric (HbO, HbR, HbT, Oxy) as shown in Figure 9.

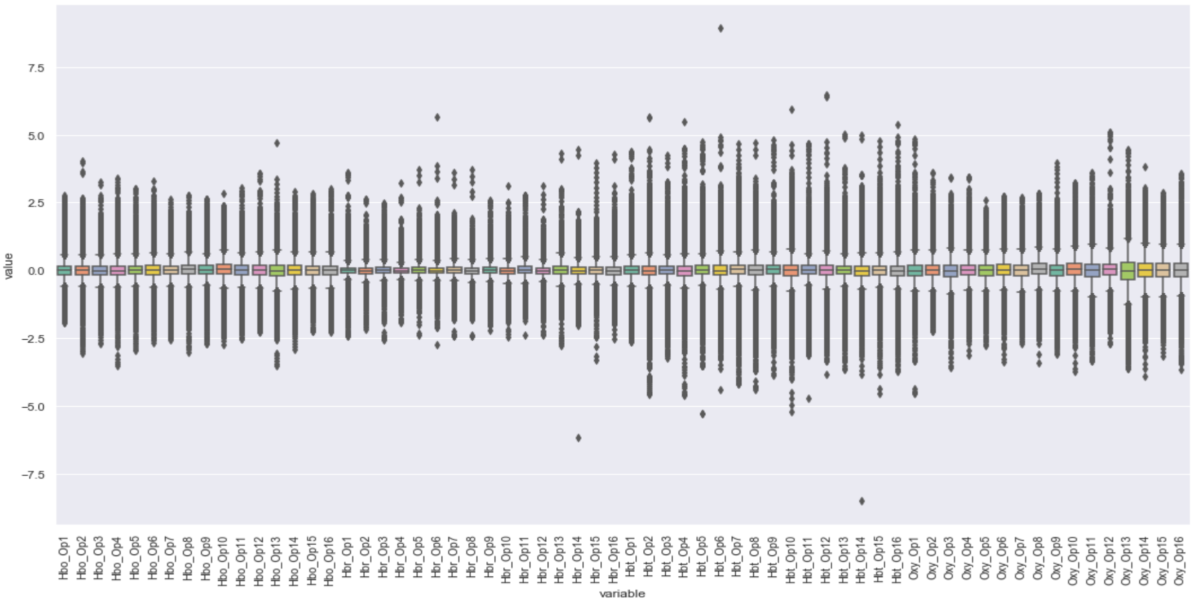

Figure 9: Boxplot of features

The mean and standard deviation of each features according to Type Code are graphed in Figure 10:

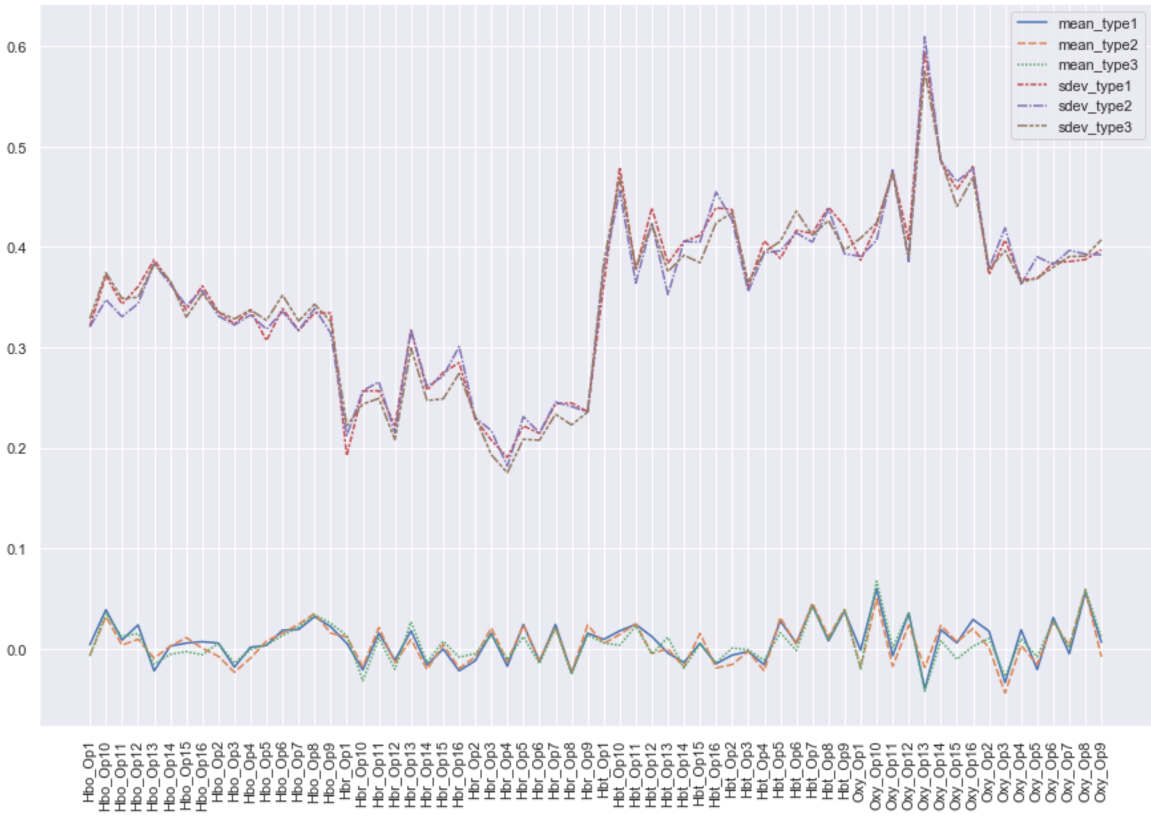

Figure 10: Mean and standard deviation of each feature according to Type Code

While the means of each feature lies around 0, standard deviations are high comparatively. Especially, the HbT and Oxy metrics have high standard deviation compared to HbO and HbR. However, mean and standard deviations look same for different Type Codes.

Similarly, density plot in Figure 11 shows that all features are concentrated around 0 and have changing kurtosis.

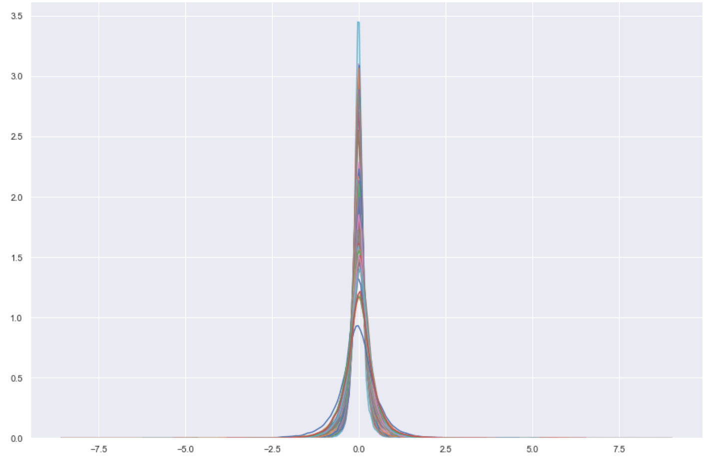

Figure 11: Density plot of features

When the correlation of features is graphed in a heatmap as in Figure 12:

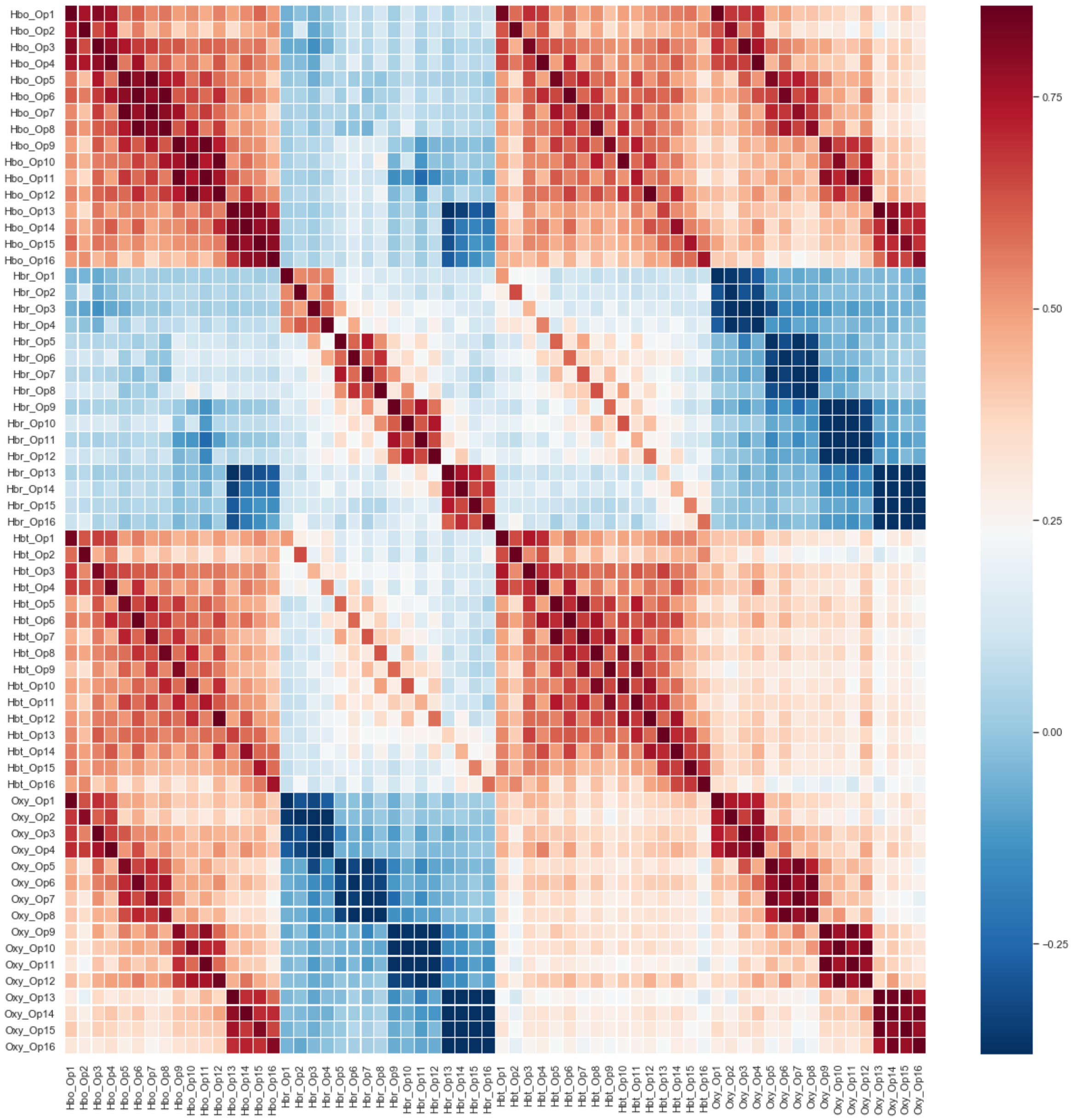

Figure 12: Heatmap for correlation of features

It is seen that;

- Optodes are grouped as 4; in other words, the first 4, second 4, third 4 and forth 4 optodes are highly positively correlated in its itself.
- HbO and HbR measurements shows relatively weak correlation
- Oxy and HbR shows highly negative correlation
- Oxy and HbO shows highly positive correlation

And the Figure 13 below shows the change in hemodynamic responses on metric HbO over the time horizon of exposure to stimuli. Each line represents an optode and each datapoint is the average of HbO values of all participants for all stimuli labeled as 1, 2 or 3.

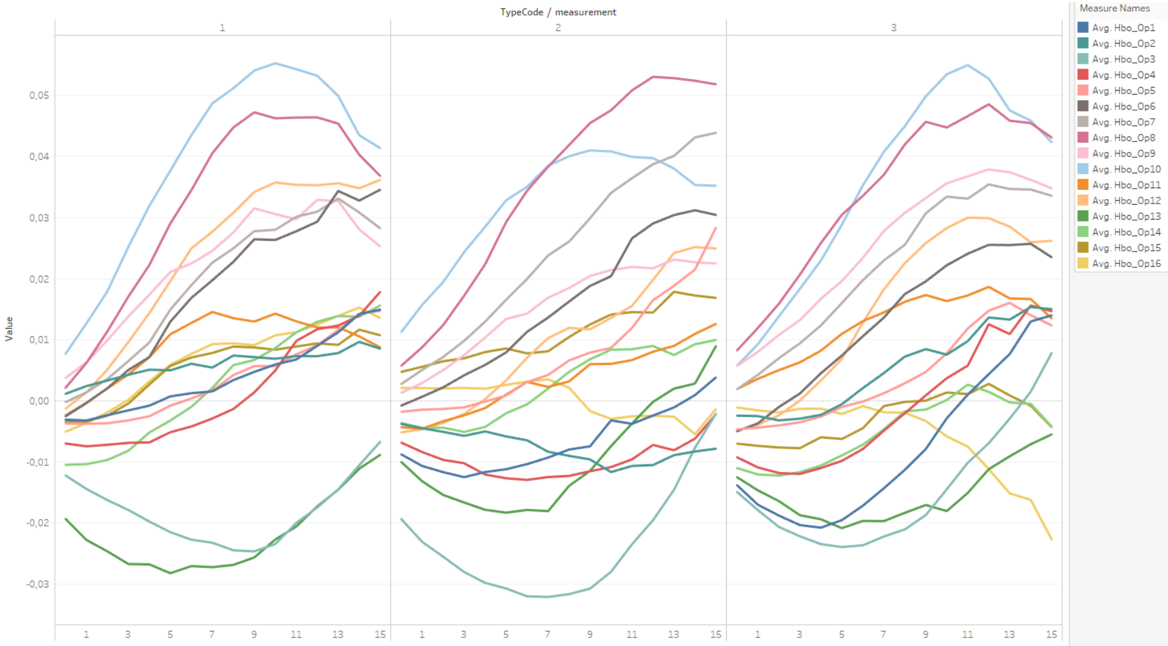

Figure 13: Change in average HbO values over time according to Type Code

The development over time seems does not have different characteristics for each Type Code. Generally, they have increasing trend with similar slope.

5. METHODOLOGY

Data Cleaning and Dealing with Null values

As the first step of cleaning, I dropped the columns "Participant", "Stimuli Number", "Procedure", "Brand", "millisecond", "Response" and "Type" since I created artificial features from these features.

The columns are checked whether they have Null values and see that every column except “Participant_Stimuli” and “measurement” have Null values. In “Response Code” column, there are 4064 rows with Null value – the ones not answered by the participant. Since the measurement values of the participants who did not answer the question are not suitable to comment on, it will be appropriate to remove these columns. For this reason, the observations whose response code is Null are removed.

After removing these observations of which “Response Code” is Null, only the first 64 rows (Hbo_Op1 to Oxy_Op16) remained with Null values. Max number of Null values is 25724 and it belongs to Optode15 for all 4 metrics. Out of 77,192 rows, the average of count of Null values is 3984 (around 5%) with standard deviation of 1707.

In order to check the Null values per Participant_Stimuli, the ratio of Null values to total number of measurements per “Participant_Stimuli” is calculated. In result, it is seen that some optodes are not take measurement either at all during 8 second of exposure or just limited measurement is taken. Since it will be not appropriate to assess hemodynamic responses with limited number of measurements, it is decided to change the Non-Null values to Null if the ratio of Null observations to all observations exceeds 32%. The reason that the ratio of %32 is decided; all measurements are taken in 8 seconds and the maximum loss of data that we can accept is 2.5 seconds, 31.25% of all time. Since there is a time lag between first exposure to stimuli and oxygenation of the tissue, the maximum loss of measurement we want to accept is 2.5 seconds.

Considering the time gap between seeing the stimuli and oxygenation-deoxygenation of tissues, 5.5 seconds are the limit that is decided to optimize cleaning as much as possible while avoiding losing meaningful data.

After changing the Non-Null values to Null, the frequency of Null values in each feature is as follows:

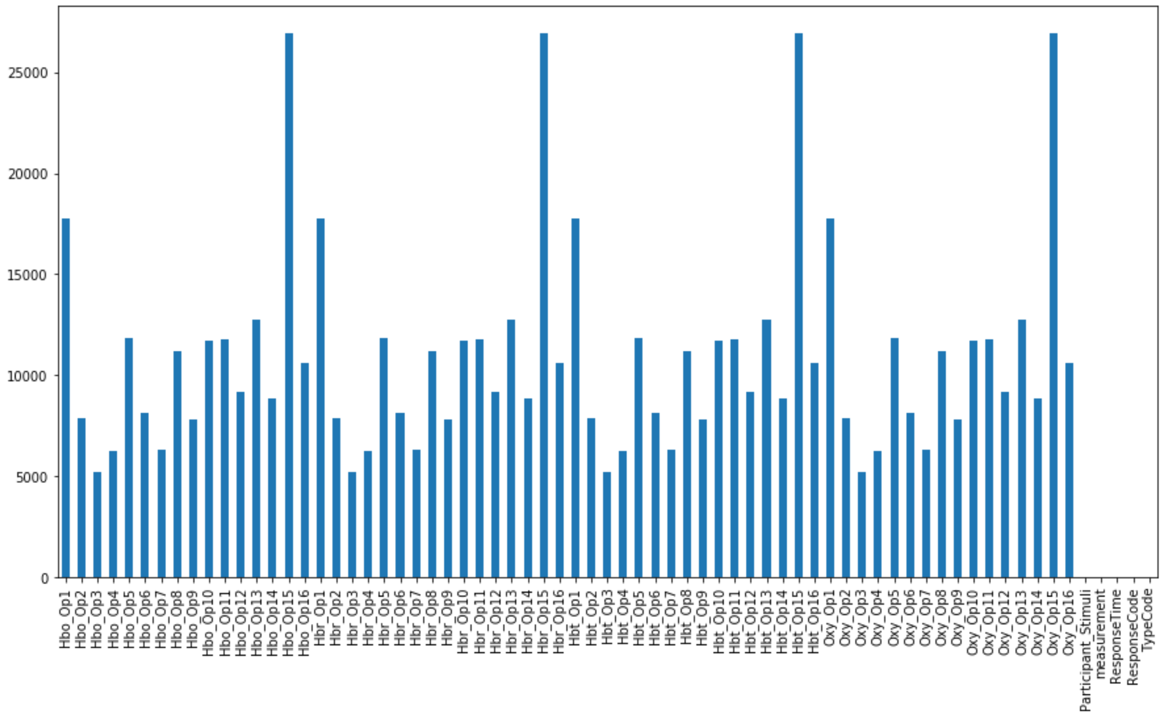

Figure 14: Frequency of Null values according to columns

Optodes 1 and 15 are the ones with maximum number of Null values, most probably it results from being at the right and left top edges of the fNIRS head band.

And the frequency of Null values according to rows are graphed below in Figure 15:

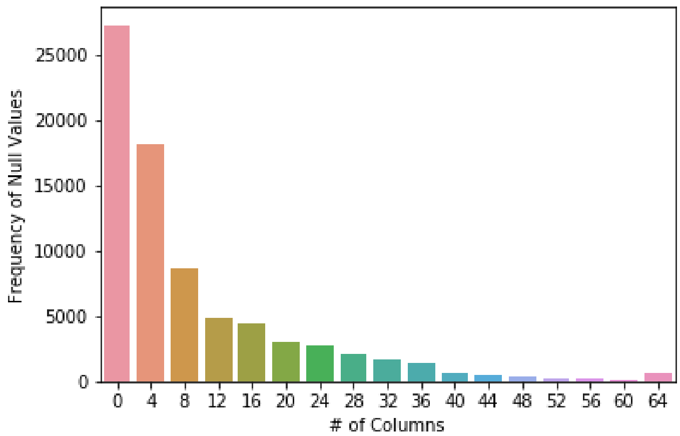

Figure 15: Frequency of Null values according to rows

As seen in Figure 15, there are more than 27,000 rows of which all the columns are complete and there are 569 rows of which all of 64 features are Null.

Considering the Figure 15, it is decided to remove the observations in case half of the optodes does not have measurement. For this reason, the observations which have Null values in more than 28 columns are removed. That means removal of total of 5,810 rows.

In result of these cleaning procedure; the data represents;

- The measurement values of the participants who answered the question,
- Measurement is taken more than 68% of the time horizon that the participant exposed to stimuli
- More than 28 features have measurement value.

The shape of the data after these cleaning steps are 71,382 rows × 69 columns.

Consolidation of Measurements – Feature Engineering

Since the dataset contains 16 or 17 measurements for each stimulus, there is need to consolidate these measurements for each participant-stimuli match. In order to transform all measurement into one metric, I decided to proceed with 2 options:

1- Taking mean of the measurements per stimuli (df_mean dataset)

2- Taking the difference of max and min points with distinguishing increasing and decreasing trends. In other words, in case min point is before max in hemodynamic response function, calculating max-min, otherwise min-max. (df_slope dataset)

For the first option, the data is grouped by with Participant-Stimuli matches and average of all measurements are taken. After that, since there are Null values for each feature, it is decided to fill the NA values with average of values of neighbor optodes as shown with blue circles in Figure16. In this technique, for the first four optodes, average of first 6; for the 5th and 6th optodes average of optode 3 to 8; for the 7th and 8th average of 5 to 10 and so and so forth, are taken. Then the technique is repeated for each 4 metrics.

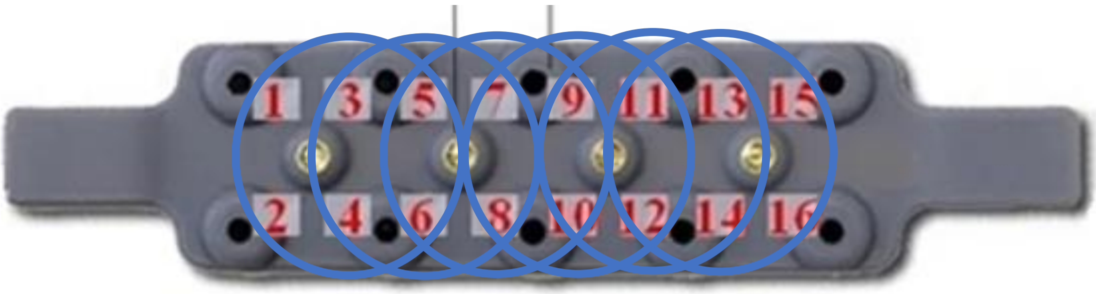

Figure 16: Neighbors of each optode used to fill Null values

After filling the Null values with 2 neighbor optode columns, I checked whether is there any Null values remaining and if there is, I dropped these rows from dataset.

And for the alternative method, I created four datasets through group by of Participant-Stimuli matches. These are:

- Minimum values of measurements
- Maximum values of measurements
- Index values of minimum values
- Index values of maximum values

Then, according to index values, if min-index is lower than max index (there is an decrease in hemodynamic response function first and then increasing later), I calculate max value minus min value, and otherwise (there is an increase in hemodynamic response function first and then decreasing later) min value minus max value. For filling the NA values, I use the same technique used in first alternative.

As seen in Figure 8 above under exploratory data analysis section, there are inconsistency between Type Code and Response Code. For this reason, I rechecked the confusion matrix of Type Code and Response Code and the result for both alternative datasets (df_mean_filtered & df_slope_filtered) are summarized in Table 4:

		Response Code		
		1	2	3
Type Code	1	1266	124	93
	2	519	357	592
	3	76	76	1344

Table 4: Confusion Matrix for Type Code and Response Code

Since I expect that the measurements with same label in Type Code and Response Code may have more reliable measurement values, in order to use more consistent dataset in machine learning algorithms and the Neutral (2) labeled stimuli are not well separated from Positive and Negative labels, I only worked with datasets filtered as both Type Code and Response Code equals to 1 or 3. So, I filtered the dataset to include 2 categories.

Since the 64 features in the dataset are 16 optodes' measurement values for 4 different metrics, which all representing hemodynamic responses, there is important correlations between them as mentioned under exploratory analytics section. For this reason, it is needed to apply dimensionality reduction for these features. Being dimensionality reduction, I applied Principal Component Analysis (PCA) to both alternative datasets.

PCA is a dimensionality reduction technique and it aims to reducing the data to its basic components, removing the unnecessary parts so the noises in the data. After calculating the principal components, we use these components as new features representing/explaining the cumulative variance in the dataset. In the calculation, eigenvector and eigenvalues are used. Eigenvector is the linear vector created using the eigenvalue which is describing how much variance there is in the data in that direction. Whenever the eigenvalue maximized with a specific eigenvector, this eigenvector becomes the principal component representing the variance of dataset. This is done basically analyzing the covariance matrix of the dataset in order to capture as much variance as possible, with the fewest component.

The cumulative variance explained with principal components (PC) for the datasets df_mean_filtered and df_slope_filtered are graphed in Figure 17 and 18:

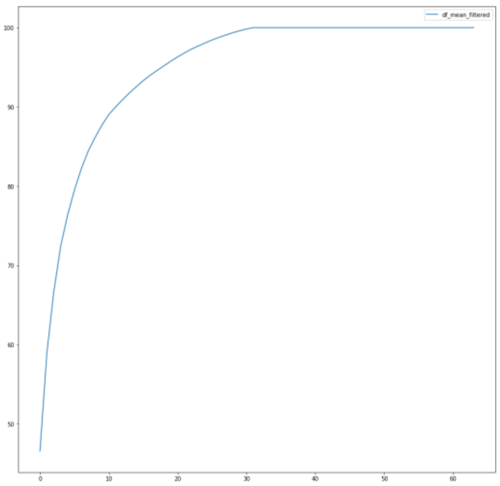

Figure 17: Cumulative Variance explained with PCs for df_mean_filtered

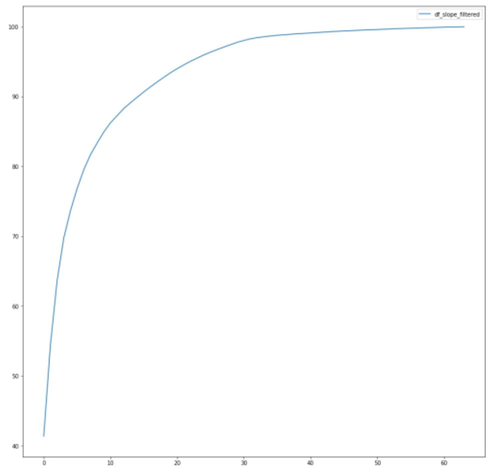

Figure 18: Cumulative Variance explained with PCs for df_slope_filtered

The cumulative variance functions look similar for both datasets. For the df_mean_filtered dataset 7 components cumulative variance exceeds 80% while it is 8 components for df_slope_filtered. And above 90% cumulative variance is reached with 12, 16 components respectively.

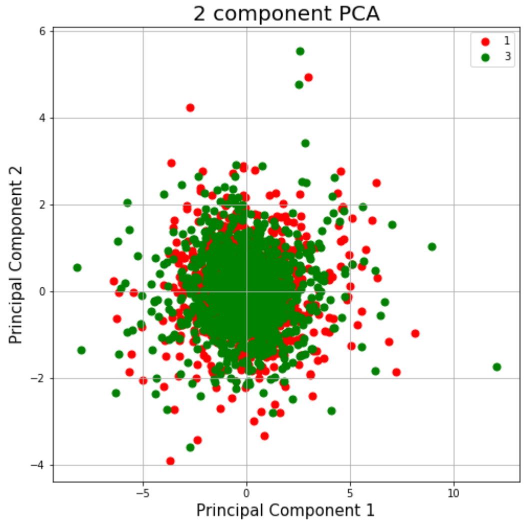

Figure 19: Scatter plot- 2 component PCA for df_mean_filtered

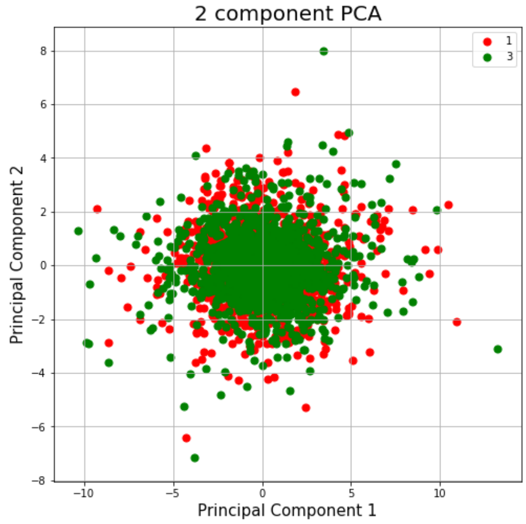

Figure 20: Scatter plot- 2 component PCA for df_slope_filtered

With the first two components, the scatter plot of datapoints are graphed above in Figure 19 and 20. They are mostly overlapped with 2 components. And when the correlations of principal components with features are checked (Figure 21 and 22), it seen that while PC0 is representing HbO, HbT, Oxy metrics, PC1 is correlated with HbR and PC2,3,4,5,6 are differentiating them from each other. PC7 and 8 are look like differentiating in optode base.

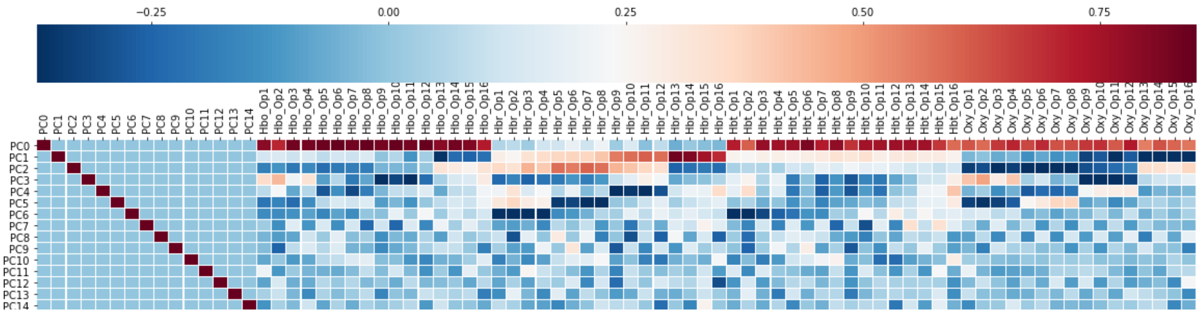

Figure 21: Correlation of Components with features for df_mean_filtered

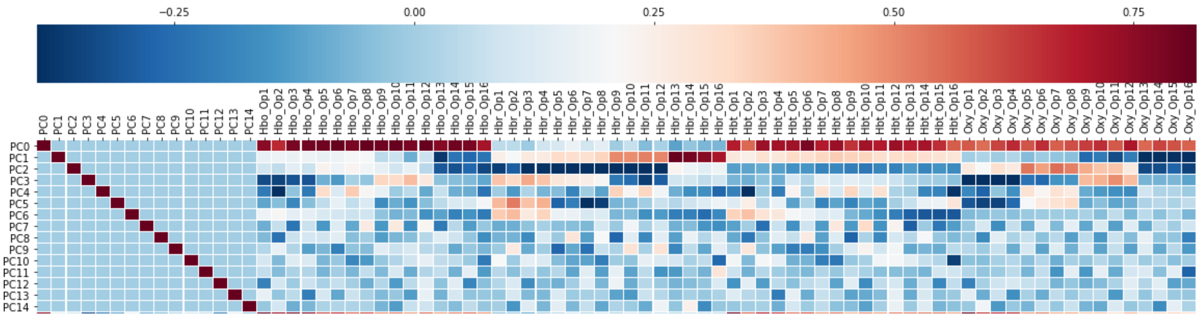

Figure 22: Correlation of Components with features for df_slope_filtered

Applying Machine Learning Models

In order to model machine learning algorithms, I use the data outputted from PCA with 6 components. And in order to train and test the models, I split the datasets 80% and 20% respectively.

Firstly, in order to evaluate rough results of classifiers, I run a set of classifiers with default parameters. The algorithms used and train and test set accuracy scores are as follows:

	df_mean_filtered		df_slope_filtered	
Classifier	**Train Set Accuracy**	**Test Set Accuracy**	**Train Set Accuracy**	**Test Set Accuracy**
KNeighbors Classifier	68.10%	49.04%	67.82%	50.57%
SVC	55.99%	49.81%	63.03%	49.62%
Decision Tree Classifier	99.86%	49.23%	99.86%	51.72%
Random Forest Classifier	98.04%	48.85%	98.04%	52.49%
AdaBoost Classifier	58.91%	46.74%	59.87%	49.23%
Gradient Boosting Classifier	74.57%	48.85%	74.62%	48.85%
Gaussian NB	52.78%	46.17%	53.02%	47.32%
Linear Discriminant Analysis	52.30%	49.04%	51.77%	50.38%
Quadratic Discriminant Analysis	52.20%	46.74%	53.93%	50.00%
Logistic Regression	52.35%	49.04%	51.72%	50.38%
Extra Trees Classifier	99.86%	45.98%	99.86%	49.04%
Bagging Classifier	98.23%	52.87%	97.75%	50.96%
XGBoost Classifier	72.27%	47.51%	73.37%	50.38%

Table 5: Rough results of classifier accuracy performance

While most of the classifier's performance is so weak, decision tree based algorithms are overfitted on train set. SVC and XGBoost Classifiers are studied on detail in following sections.

On df_mean_filtered dataset running the SVC algorithm;

With kernel type of Radial Basis Function (rbf) and penalty parameter of the error term (C) of 1, train and test accuracies are 55% and 49% respectively. The C parameter explains how much tolerance is given to misclassification; the bigger C value gets lower tolerance. And the kernel rbf is the method to draw non-linear decision boundary by calculating and maximizing the distance between each data point to all other points while the kernel poly applies polynomial combination of existing features in order to draw non-linear decision boundary. Considering the Figure 19 and 20, the decision boundary should be non-linear. Through training by changing the "C" between 0.1 and 50 with iteration of

0.3 and for kernel “rbf” and “poly” (with 3rd degree), test set accuracies do not get better performance. (Figure 23 and 24) Increasing the “C” parameter in “rbf” kernel tends to increase overfitting, while does not make significant effect in kernel “poly”.

And in case the degree changes in kernel “poly”, increasing degree increases the accuracy on train set but does not affect significantly on test set which is also commented as overfitting.

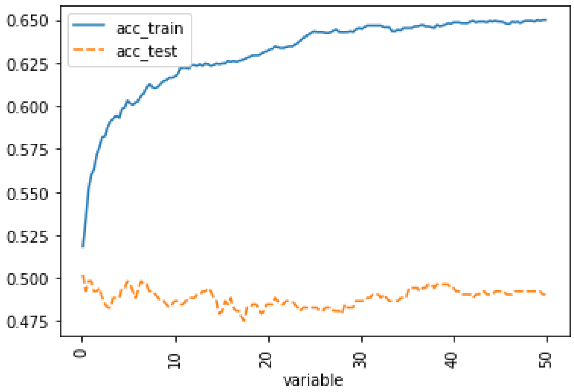

Figure 23: Accuracy vs “C” parameter (df_mean_filtered with kernel “rbf”)

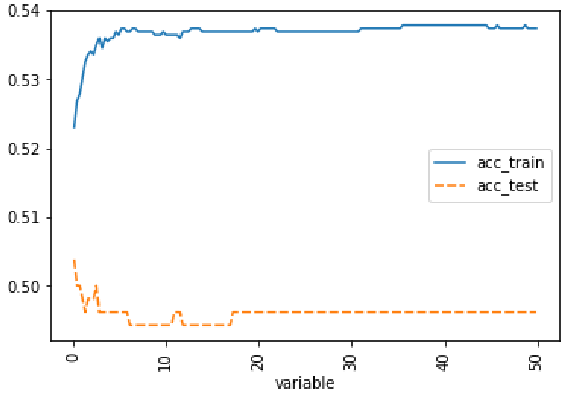

Figure 24: Accuracy vs “C” parameter (df_mean_filtered with kernel “poly”)

On df_slope_filtered dataset;

With the same parameters used in df_mean_filtered dataset, train and test accuracies are 63% and 49% respectively. Like the df_mean_filtered dataset, changing the “C” value on rbf and poly kernels, keeping other parameters constant, although the accuracy on training dataset increases with increase in “C”, it is not increasing the accuracy of test dataset. (Figure 25 and 26)

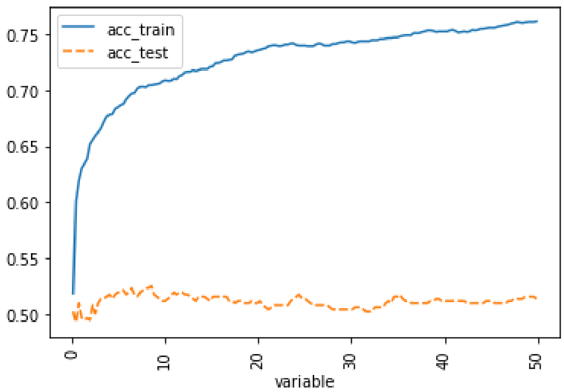

Figure 25: Accuracy vs “C” parameter (df_slope_filtered with kernel “rbf”)

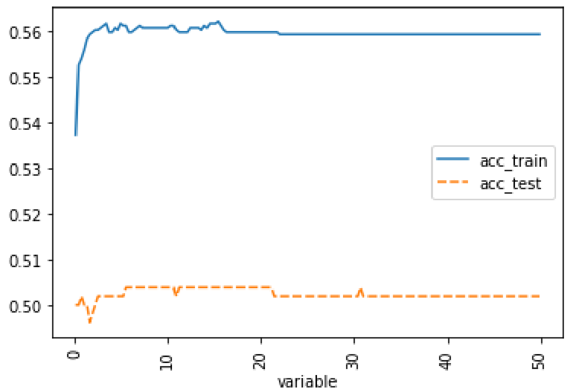

Figure 26: Accuracy vs “C” parameter (df_slope_filtered with kernel “poly”)

I also, run the SVC algorithm for both datasets in grid search with various alternative parameter values however could not get the better model performance.

Second algorithm used in this project is XGBoost Classifier. XGBoost is an ensemble machine learning model combining several base models in order to produce one optimal predictive model and it uses decision tree ensembles.

With using learning rate of 1 and max depth of 5, train and test accuracies are around 99% and 50% respectively in both datasets which is highly overfitted. While the learning rate parameter controls how big iteration on each step to find the label, max depth parameter defines the depth of the tree as number of levels. In order to prevent overfitting, firstly, the learning rate was changed. When the learning rate exceeds 2, both training and test set accuracies are being stable around 50%. (Figure 27)

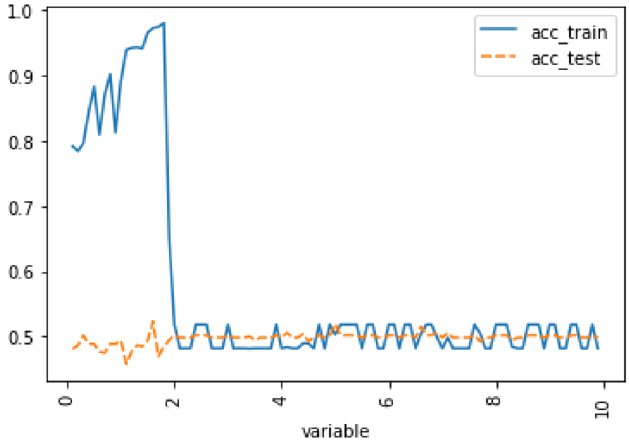

Figure 27: Accuracy vs "learning rate" parameter (df_mean_filtered)

When learning rate set as 2, accuracy changes as figured below in Figure 28 with changing max_depth parameter. After 6, it overfits on training set. And below 6, both accuracies are around 50% which is very low performance.

Genele Açık

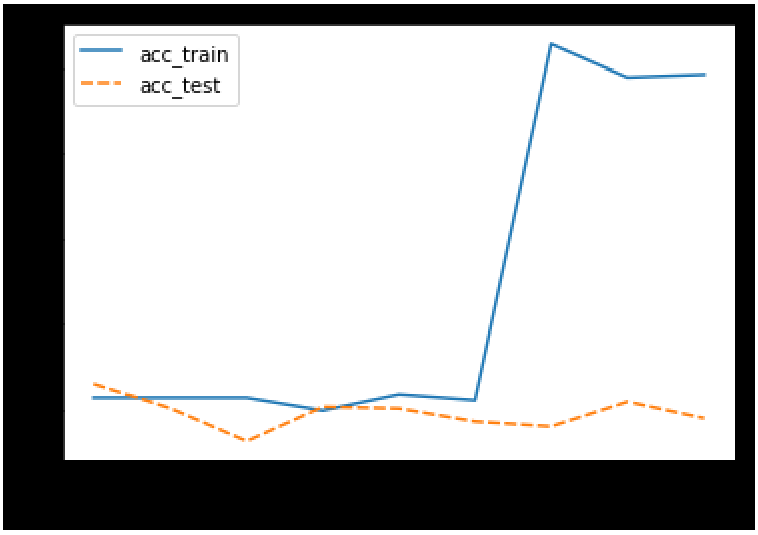

Figure 28: Accuracy vs "max depth" parameter (df_mean_filtered)

When the same parameters are used in df_slope_filtered dataset, the performance of model is similar with df_mean_filtered dataset.

With grid search, the parameters in Table 7 are used to tune the parameters of XG Boost Classifier algorithm.

Parameter	**Variable**
learning rate	[1,2,3,4,5]
n_estimators	[10,100,200]
max_depth	[2,3,4,5,10,15]
min_samples_split	[1,10,50,100]
min_child_weight	[2]
fold	5

Table 6: Parameters used in Grid Search for XGBoost

The accuracy score is 51% with best parameters set. The accuracy is similar when the same parameters are used in df_slope_filtered dataset.

Besides, the df_mean_filtered dataset is studied on Microsoft Azure Machine Learning Studio with the following algorithms:

- Two Class Averaged Perceptron
- Two Class Bayes Point Machine
- Two Class Boosted Decision Tree
- Two Class Decision Forest
- Two Class Locally Deep Support Vector Machine
- Two Class Support Vector Machine
- Two Class Logistic Regression

And again, the accuracies for all the algorithms are around 50% which is showing the model is unsatisfactory. Besides, with two class neural network algorithm and tuning the parameters with entire grid mode, still the model remained unsatisfactory with accuracy level of slightly over 50%.

6. CONCLUSION AND FURTHER STUDIES

In order to predict the consumer perception on brands, I studied on hemodynamic responses of a group of experiment participants recorded with functional near-infrared spectroscopy (fNIRS) device. In order to review the similar studies performed formerly, first literature review was performed. In data preparation phase, after reading and consolidating the source data taken from fNIRS device, the dataframe tidied up. Following the descriptive and exploratory analysis of data where the first insights are taken about the data, the methodology to clean the data, selecting and designing the features are applied. Lastly, machine learning models are built, and results are evaluated.

After running a set of classifiers with default parameters, performances of classifiers are very low, slightly over 50% while decision-tree based algorithms overfitted with train dataset. With support vector classifier and extreme gradient booster algorithms, further studies are performed to optimize models. However, with different parameters used in these algorithms, better performance was not accomplished.

Since the mean and difference of max and min points in hemodynamic response function were used as representative of measurement taken from each participant-stimuli match, further studies should be focused to perform feature engineering studies with different options. The options should extinguish the development of hemodynamic response function for each class namely Positively Related, Negatively Related and Neutral.

7. REFERENCES AND CITING

of purchase behavior through fNIRS. European Journal of Marketing, 224-243.

Gervain, J., Mehler, J., Werker, J., Nelson, C., & Csibra, G. (2011). Near-infrared spectroscopy: A report from the McDonnell infant. Developmental Cognitive Neuroscience, 22-46.

Hosseini, S. H., Mano, Y., Rostami, M., Takahashi, M., Sugiura, M., & Kawashima, R. (2011). Decoding what one likes or dislikes from single-trial fNIRS measurements. NeuroReport, 269-273.

Kopton, I. M., & Kenning, P. (2014). Near-infrared spectroscopy (NIRS) as a new tool for neuroeconomic research. Frontiers in Human Neuroscience, 8.

Krampe, C., Strelow, E., Haas, A., & Kenning, P. (2018). The application of mobile fNIRS to first insights from a merchandising communication study. European Journal of Marketing

Liu, X., Kim, C.-S., & Hong, K.-S. (2018). An fNIRS-based investigation of visual merchandising displays for fashion stores. PLoS ONE

Quaresima, V., & Ferrari, M. (2016). Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise Review. Organizational Research Methods

Santos , J., Seixas, D., Brandão, S., & Moutinho , L. (2012). Neuroscience in branding: A functional magnetic resonance imaging study on brands implicit and explicit impressions. Journal of Brand Management, 735 757.

Sebastian, V. (2014). New directions in understanding the decision-making process: neuroeconomics and neuromarketing. 758 762.

TABLE OF FIGURES

TABLE OF TABLES