MINING EYETRACKING DATA TO CHARACTERISE USERS AND THEIR PATTERNS OF USE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

MELİH ÖDER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JUNE 2019

Approval of the thesis:

MINING EYETRACKING DATA TO CHARACTERISE USERS AND THEIR PATTERNS OF USE

Submitted by **MELİH ÖDER** in partial fulfillment of the requirements for the degree of **Master of Science in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, Graduate School of **Informatics** ______________

Prof. Dr. Yasemin Yardımcı Çetin
Head of Department, **Information Systems** ______________

Assoc. Prof. Dr. Aysu Betin Can
Supervisor, **Information Systems Dept., METU** ______________

Assoc. Prof. Dr. Yeliz Yeşilada
Co-supervisor, **Computer Engineering Dept., METU NCC** ______________

Examining Committee Members:

Assoc. Prof. Dr. Banu Günel Kılıç
Information Systems Dept., METU ______________

Assoc. Prof. Dr. Aysu Betin Can
Information Systems Dept., METU ______________

Assoc. Prof. Dr. Tuğba Taşkaya Temizel
Information Systems Dept., METU ______________

Assoc. Prof. Dr. Nergiz Çağıltay
Software Engineering Dept., Atılım University ______________

Assist. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MELİH ÖDER

Signature :

ABSTRACT

MINING EYETRACKING DATA TO CHARACTERISE USERS AND THEIR PATTERNS OF USE

ÖDER, MELİH
M.S., Department of Information Systems
Supervisor : Assoc. Prof. Dr. Aysu Betin Can
Co-Supervisor : Assoc. Prof. Dr. Yeliz Yeşilada

June 2019, 88 pages

Eye tracking studies typically collect an enormous amount of data that encodes a lot of information about the users' behavior and characteristics on the web. However, there are not many studies that mine such data to learn and discover user characteristics and profiles. The main goal of this study is to mine eye tracking data by machine learning methods to create data models which characterise users and predict their characteristics, in particular, familiarity and gender. Detecting users' characteristics can be used in creating adaptive user interfaces to improve user experience and interaction efficiency. In a typical eye tracking study, collected demographics data have participants' educational backgrounds, gender, age, and frequency of the web page use. In this thesis, a model focusing on the users' familiarity degree and gender is first created based on an existing eye-tracking dataset, and then a new eye-tracking study is conducted to validate this model. The main contribution of this thesis is a machine learning approach that can be used to characterise users, in particular, familiarity and gender, based on eye-tracking data and also a tool that can be used to extract features and metrics from an eye-tracking dataset.

Keywords: Eye tracking, user modelling, data mining, familiarity, gender

ÖZ

VERİ MADENCİLİĞİ YÖNTEMİYLE GÖZ İZLEME VERİLERİNİ İŞLEYEREK KULLANICILARI VE KULLANIM YÖNTEMLERİNİ KARAKTERİZE ETME

ÖDER, MELİH
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Yöneticisi : Doç. Dr. Aysu Betin Can
Ortak Tez Yöneticisi : Doç. Dr. Yeliz Yeşilada

Haziran 2019 , 88 sayfa

Göz izleme çalışmalarında toplanan veriler, kullanıcıların web üzerindeki davranışları ve karakterleri hakkında bol miktarda bilgi içermesine rağmen bu verileri işleyerek kullanıcı profillerini tahmin etmeye çalışan çok fazla çalışma bulunmamaktadır. Bu çalışmanın ana amacı, göz izleme verilerini makine öğrenmesi yöntemleriyle işleyerek kullanıcı özelliklerini karakterize eden veri modelleri çıkarmak ve özellikle kullanıcıların web sayfasına aşinalığını ve cinsiyetlerini tahmin etmeye çalışmaktır. Kullanıcı özelliklerini tahmin etmek, adaptif web sayfaları tasarlayarak kullanım kolaylığı sağlamaya yarayabilir. Göz izleme çalışması sırasında, kullanıcıların eğitim geçmişleri, cinsiyetleri, yaşları ve çalışma sırasında kullanılan web sayfalarını kullanım sıklıkları sorulmuştur. Bu çalışma sırasında, öncelikle var olan veri seti kullanılarak veri modelleri çıkartıldı ve tekrardan göz izleme çalışması yapılarak, veri modelleri doğrulandı. Çalışmanın, göz izleme verilerini makine öğrenme yöntemleriyle işleyerek, kullanıcıların web sayfalarına aşinalığını ve cinsiyetlerini karakterize etmesi ve eğitilecek veri setini hazırlayan bir araç geliştirmiş olması literatüre sağladığı katkılardır.

Anahtar Kelimeler: Göz izleme, kullanıcı modelleme, veri madenciliği, aşinalık, cinsiyet

To my lovely wife

ACKNOWLEDGMENTS

I would like to express my special thanks of gratitude to my supervisor Yeliz Yeşilada who gave me the golden opportunity, support, and encouragement to complete this project. Her valuable feedbacks led me in terms of both content and method of the thesis.

I also would like to thank my supervisor Aysu Betin Can who enthusiastically guided me in administrative processes and facilitateted my work as well as she can.

I also would like to acknowledge Pınar Karagöz who gave valuable feedbacks to me at the beginning of this study which provided the thesis with a direction.

Besides my teachers, I would like to specially thank Dr. Şükrü Eraslan who shared his knowledge and documents about eye tracking studies. Moreover, he supported me with his valuable feedbacks.

Lastly, I dedicate my thesis to my lovely wife (Sümeyye Öder) who have encouraged and motivated me to complete the thesis without any problem. I also dedicate my thesis to my mother and father have trusted and supported me throughout my education. I would like to give a special thanks to my grandmother who has lived with and looked after me with a deepest sacrifice in my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AOI	Area of Interest
DBSCAN	Density-based spatial clustering of applications with noise
SVM	Support Vector Machines
PCA	Principal Component Analysis
SOM	Self Organizing Map
LDA	Linear Discriminant Analysis
SB	Sequence-based
PB	Page-based
ITS	Intelligent Tutoring System
SMO	Sequential Minimal Optimization
QP	Quadratic Programming
FE	Feature Extraction
VIPS	Vision-based Page Segmentation
ID	Identification
IDE	Integrated Development Environment
MS	Microsoft
MS	Milliseconds
PXL	Pixel
STA	Scanpath Trend Analysis
SD	Standard Deviation
METU NCC	Middle East Technical University Northern Cyprus Campus
KNN	K-nearest neighbors
SMOTE	Synthetic Minority Oversampling Technique
LR	Logistic Regression

CHAPTER 1

INTRODUCTION

The web plays an important role in our daily lives. In order to attract more users and make them revisit web pages, user-friendly design is important as much as the content. Eye-tracking studies have been widely used to assess web design and in particular their usability to enhance them to user-friendly designs [3]. In most of these studies, a web usability specialist creates sample scenarios over a specific site which needs to represent usage profiles of that site. Then, users are invited to implement specific tasks on that site.

After that, in ordinary web usability tests, the usability specialist evaluates whether objectives of the scenario are achieved or not. While evaluating, a checklist is generally used which is prepared according to the scenario. Furthermore, outputs of the eye tracker may be used to show some specific achievements. Gaze points map and heatmap are mostly used outputs to understand users' eye movements and interests. Based on the usability specialist's evaluation and report, web design may be improved to be more user-friendly. This is the most common purpose of conducting eye-tracking studies in Turkey[4].

This work differs from ordinary eye-tracking studies because it is not conducted for a usability test; however, it is conducted to model users' familiarity and gender and predict them from eye movement data. This study intends to create eye-tracking data models by data mining techniques in order to determine if users are familiar to a web page and classify users simply based on their gender as male or female. Determining a user's familiarity and gender from their eye movement data could enable us to the design of more user-friendly pages based on user's familiarity and gender and even adapt pages based their profiles to better meet their needs.

In this thesis, first we would like to clarify what I mean by familiarity and also the scope of gender classification. When I say familiarity, actually it is referred to as familiarity to a web page as "close acquaintance with or knowledge of a particular web page"[1]. According to the definition, familiarity implies a degree of knowledge.

[1] Adopted from Oxford Dictionary definition: https://en.oxforddictionaries.com/

In eye-tracking studies, for each web page, it is asked to the participants as "how often do you visit the web site?" and they answer this question by choosing a number from an ordinal scale from 1 to 5. By the way, it is intended to decrease the subjectivity of their familiarity. This thesis intends to classify the participants based on their familiarity and gender.

1.1 Proposed Method and Models

Eye-tracking data consists of a sequence of fixations and saccades which are detected and recorded by eye-tracking equipment. It is possible to make inferences about user perception by analyzing those fixations and saccades. There are some papers which show that there are variations in eye-tracking metrics between different familiarity and also gender [5, 6, 7, 2]. Furthermore, other studies show that fixations and saccades of a user construct a complex pattern which can be detected and analyzed by data mining techniques [8, 9].

In the modeling study, for each web page, existing eye-tracking data are trained separately and data models are created. Each participant looks at each web page twice; browsing and searching purposes. Data models are also created for each purpose separately. Furthermore, each dataset is trained by two data mining techniques; Logistic Regression and Support Vector Machines. Besides the raw dataset, each dataset is preprocessed by both resampling and oversampling to smooth datasets. Therefore, there are 3 web pages and 36 data models for familiarity in total. Moreover, there are 6 web pages and 72 data models for gender in total and each data model is evaluated separately.

In validation study, eye-tracking study is conducted again to collect new eye movement data and another 20 participants participated in this study. These participants' eye movement data is used to validate extracted data models of modeling study. This validation is conducted in two ways. Firstly, new eye movement data is added to the existing one and re-train datasets to create data models with more instances. It is expected that their 10-fold cross-validation results are better than the data models of modeling study. Secondly, by using the data models of modeling study, familiarity and gender factors of the new eye movement data are predicted. Thus, it is expected that prediction results need to be as high as 10-fold cross-validation modeling results.

1.2 Contributions

The contributions of the thesis are two-fold. Firstly, the eye movement data is trained by data mining techniques and modeled to detect user's familiarity and gender which

---

definition/familiarity.

could lead to user-adaptive web designs. This method is not unique, but its approach is new. The proposed approach is Predicting user's familiarity and gender from their eye movement data in order to cause user-adaptive web design. Secondly, in order to prepare an eye-tracking dataset from raw eye movement data, a tool is developed and published as open-source which can be utilized and enhanced by other researchers. This is another contribution to the literature. In the scope of the thesis, research questions are constructed based on related work. Although they are presented in this chapter, after completing the modeling and validation of the data models, in conclusion, they will be discussed in detail. There are two research questions of this thesis.

Research Questions;

- Can a familiar user to a web page predicted from the user's eye-tracking data by using data mining techniques?
- Can a user predicted as male or female from the user's eye-tracking data by using data mining techniques?

1.3 The Outline of the Thesis

This thesis comprises of six main chapters; Introduction, Related Work, Feature Extraction Tool, Modeling Study, Validation Study, and Conclusion. Firstly, Introduction Chapter introduced the thesis in terms of its purpose, methods, and contributions. Then, in Related Work Chapter, purposes, methods, and, outputs of existing eye-tracking studies are examined in order to show the gap in the literature. Moreover, data mining techniques are examined in order to detect the most suitable one for binary classification. Furthermore, familiarity and gender-related gaze features are detected to create datasets for classification.

Raw eye movement dataset needs to be prepared for training. Feature Extraction Tool is constructed to organize eye movement data under certain features and Feature Extraction Tool Chapter describes this tool's architecture and usage. Modeling Chapter explains the methodology to extract data models. In order to enhance data models' quality, preprocessing and feature selection need to be conducted which will be also explained in detail. Moreover, accuracy, recall, precision, and F-measure values are provided as the modeling results in tables.

In the Validation Chapter, an eye-tracking study is conducted again. Moreover, this chapter explains the details of this new eye-tracking study and how to validate the existing data models. Lastly, the Conclusion Chapter will discuss modeling, validation results and the research questions in detail. It will show the limitations and future work of this thesis.

CHAPTER 2

RELATED WORK

This work aims to predict users' familiarity to the web pages and genders by mining their eye movement data. This may boost web designs to create adaptive web pages to address the right users. As a result, this chapter reviews existing related-studies under three sections; eye tracking studies, data mining techniques, and gaze features. Moreover, this section presents the gap in the literature and shows the reason why this research is conducted.

2.1 Eye Tracking

Resources show that there are different purposes of eye tracking studies although mostly the eye movements are utilized for usability tests. Firstly, in the earlier time of eye tracking, Loftus and Mackworth [10] conducted an experiment which aimed to understand the cognitive determinants of viewing a picture. In 1978, they recorded participants eye movements by a camera. This study revealed that informative objects attract more attention than non-informative ones. Therefore, familiar users can probably know where informative elements of a web page; so, their eye movements are different than unfamiliar ones.

Secondly, Rayner [11] introduced that while reading and processing information, eye movements are differed with respect to reader's velocity, age, and dyslexia. Rello, and Ballesteros [8] firstly modeled user's eye movements to detect if they are dyslexic or not. Moreover, in 2018, a similar study was conducted for autism and the results are very promising [9]. Furthermore, the familiarity effect on eye movements researched by Greene and Rayner [5] at the year of 2001. They conducted four experiments and then showed that familiar users have longer and fewer fixations over a web page. These studies show that eye tracking data is not just random figures; in contrast, it may be a good indicator of user's characteristics if it is analyzed and mined.

In addition, at 2004, Pan examined the determinants of the eye movement behavior who identified gender, viewing order and web site type as determinants [7]. An eye tracking study was conducted to collect eye movement data and then they converted data to the scanpaths and analyzed by String-Edit Method which is introduced by Josephson and Holmes at 2002 [12]. Based upon scanpath analysis, reengineering web pages for constraint environments such as visually disabled users was researched and implemented [13]. And even a scanpath analysis algorithm was created to construct a common scanpath for similar users and facilitate to classify users with respect to their scanpaths [1].

Table 2.1: Eye-Tracking Related Work

Ref	Purpose	Technique	Sample Size (female + male)	Features Related				
				Fixation Duration	Path Angles	Fixation Counts	Fixation Distance	Predefined AOIs
[6]	Investigating relationship between visual memory and gaze features	- DBSCAN (clustering) - Permutation test (non-parametric test)	24 subjects (10f + 14m)	✔	✔	✔	✗	✗
[8]	Identifying if a user is dyslexic or not	Support Vector Machine (SVM)	97 subjects (50f + 47m)	✔	✗	✔	✗	✔
[9]	Identifying if a user is autistic or not	Logistic regression	30 subjects	✔	✗	✔	✗	✔
[14]	Relevance of document titles to search tasks	-PCA -SOM -Linear discriminant analysis (LDA)	3 subjects	✔	✗	✔	✔	✗
[15]	Clustering eye tracking recordings as representation of viewer interest	Mean shift procedure	6 subjects	✔	✗	✗	✔	✗
[16]	Assessing student learning	Simple logistic regression	47 subjects	✔	✔	✔	✔	✔
[17]	Identifying behavioural patterns of use	-Differential sequence analysis -PCA	-	✗	✗	✗	✗	✔
[18]	Designing information visualisation systems dynamically adapt to user characteristics	-Statistical analysis -PCA	35 subjects (18f + 17m)	✔	✔	✔	✔	✔

Table 2.1 shows related eye tracking studies which are trained by data mining techniques. Each of these has different purposes; but Rello [8], Yaneva [9] and Bondareva [16]'s studies are similar to this thesis in terms of outputs. All of them classify the data to infer a binary output. In addition, this table indicates data mining techniques which can be classification or clustering method. In order to train eye movement data, it needs to determine eye tracking features such as fixation counts, fixation duration and so on which need to represent eye movement data correctly. Table 2.1 shows mostly used eye tracking features; fixation duration, fixation counts, fixation distance (saccade length), path angles (between consecutive fixations) and predefined AOIs. The next section will discuss about the data mining techniques used in the related work and we discuss which method will be the right method used in our study and then the right classification method will be defined for this work.

2.2 Familiarity and Gender - Gaze Features

Gaze features have been utilized for not only data mining research but also different research purposes. In fact, it is possible to infer how eye movements behave by monitoring gaze features. In this section, according to related work, gaze features are examined and familiarity and gender-related 16 gaze features are extracted. In Tables 2.2 and 2.3, 16 gaze features are summarized. We investigate familiarity-related and gender-related hypotheses to determine which can be used to differentiate user characteristics.

While investigating the relationship between visual memory and eye-tracking features, Marchal [6] uses fixation duration, fixation counts and path angles. Path angle means an angle between two consecutive fixations with respect to the +x axis as zero degrees. In Salojarvi et al.'s study, to infer implicit feedback from eye movement data, fixation duration, fixation counts and fixation distance which is a distance between two consecutive fixations are utilized [14]. While clustering eye movement data for characterizing viewer's interest, at 2004, Santella and DeCarlo also utilize fixation duration and fixation distance [15]. In addition, Steichen et al. uses predefined AOIs in their study which attempts to create a user adaptive information visualization system [17]. Also another information visualization study, Toker et al.'s study, exploits almost all eye tracking features; fixation duration, fixation counts, fixation distance, path angles, and AOIs [18]. In his study, Bondareva et al [16] mainly separated gaze features into AOI-based and non-AOI features. As seen in, Sequence-based (SB) and Page-based (PB) features represent AOI-based and non-AOI features, respectively.

Scanpath and Fixation Duration are the most commonly used metrics in the studies; however, First Fixated AOI was not used in any study and we expected that it would be a significant metric to imply user's familiarity and gender. Besides, fixation counts, saccade length (distances between fixations) and path angle variables are added to this study because various metrics might make a difference between familiar and unfamiliar users, similarly male and female [19, 6]. On the other hand, Bondareva et al. [16] stated that a large number of features may result in over-fitting data models which do produce corrupted consequences. In order to overcome this issue, in our work, we explore techniques to do feature selection.

There are familiarity and gender-related hypotheses; in that, which gaze features indicate familiarity and gender characteristics of the users (*see Tables* 2.2 and 2.3). We have reviewed similar studies and noted the hypotheses in the tables. From the familiarity-related hypotheses 2.2, firstly, fixation duration shows task difficulty and information complexity [11]. Secondly, the scanpath length is changed in terms of familiarity. Eraslan and Yesilada stated that the length of common Scanpath is equal to 2.67 as the average for familiar users, while it is 1.67 for unfamiliar group [20]. Moreover, users are inclined to fixate more to unfamiliar distractors; but less to familiar ones [5]. In other words, fixation counts and fixation counts over AOI might imply familiarity. Lastly, about the path angle between fixations, Marchal et al. stated that path angle is a very good indicator of familiarity [6]. Therefore, related works may show that eye gaze features could predict and imply user's familiarity level to a web page. It means that this is valuable to be investigated.

In addition to the familiarity-related hypotheses, eye gaze features are hypothesized for gender in Table 2.3. Firstly, Pan et al. express that females are more focusing on a comprehensive process of information while males are keeping their attention and looking to a fewer number of areas [7]. Under the illumination of this investigation, it can be hypothesized that females look longer than males; moreover, females make more fixations than males. In this section, we have noticed that there are many eye-tracking studies which reveal gaze features are indicators of the user's familiarity and gender; however, there is no work that has tried to use gaze features to predict users' familiarity or gender.

2.3 Data Mining Technique

Data mining techniques have generally been used to detect patterns in order to understand them and enhance their purpose of use. This research intends to take advantage of these techniques; but, it needs to review the past usages from Table 2.1 to determine the most accurate method.

By taking classification techniques into consideration, firstly in 2013, Bondavera et al.'s study [16] attempts to create an Intelligent Tutoring System (ITS) which aims at assessing student's learning. In fact, it trains data models according to high and low learners' eye movements. Then, the system classifies students by evaluating their eye movement data. Its output is a binary factor and resemble to this thesis. Moreover, Bondavera obtains the best data models by training with Simple Logistic Regression.

Secondly, in 2015, Rello et al. [8] by using a classification algorithm; Support Vector Machines, classifies user's eye movement data to infer if the one is dyslexic or not which is also a binary factor. Similarly, in 2018, Yaneva et al. implements (Multiple) Logistic Regression method to classify users as autistic or not. Those data models are very promising and valuable; so, in this thesis, it is decided to utilize the same supervised data mining techniques to classify a binary factor successfully. However, they need to be investigated to determine the most suitable and the best one.

According to the explanations of McDonald's book [25], Simple Logistic Regression is suitable for the one dichotomous outcome (dependent) with one independent variable although Multiple Logistic Regression differently means multiple independent variables [26]. In this case, there are various independent eye tracking features; so (Multiple) Logistic Regression looks more appropriate.

Table 2.2: Feature-Level Related Work and Familiarity-Related Hypothesis

Type	Feature	How to Compute	Familiarity-Related Hypothesis
SB	Scanpath	Shows the sequence of AOIs that a user looks one by one.	There are differences in scanpaths of familiar and unfamiliar user [5, 6, 20, 21].
SB	Mean of Sequence based Fixation Durations	Mean of fixation durations over AOIs.	There are differences in fixation durations between familiar and unfamiliar user [7, 21, 3].
SB	Sum of Sequence based Fixation Durations	Sum of fixation durations over AOIs.	Familiar user's fixation duration is longer than unfamiliar one [11].
SB	Sequence based Fixation Counts	Number of fixations over AOIs.	Familiar user makes fewer fixations than unfamiliar user [11, 5, 6].
SB	First Fixated AOI	AOI that the participant looks at first.	First Fixated AOI differs in terms of familiarity.
SB	Percentage of First Fixated AOI	Percentage of the first fixated duration within whole duration.	Duration of First Fixated AOI is different than other fixations as a percentage in terms of familiarity.
SB	Duration of First Fixated AOI	Duration when the participant looks at the first AOI.	Familiar user's fixation duration on First Fixated AOI is longer than unfamiliar one [11].
PB	Mean of Page based Fixation Durations	Mean of fixation durations.	Familiar user's fixation duration is longer than unfamiliar one [11].
PB	Sum of Page based Fixation Durations	Sum of all fixation durations.	Familiar user's fixation duration is longer than unfamiliar one [11].
PB	Page based Fixation Counts	Count all fixations	Familiar user makes fewer fixations than unfamiliar user [11, 5, 6].
PB	Number of Viewed AOIs per Page based Fixations	Fragment of number of fixations over AOIs per Number of all fixations	There are differences between familiar and unfamiliar user [21, 22].
PB	Mean of Distances among Page based Fixations	Calculates the average of distances among all points.	Fixation distance is longer for unfamiliar user than familiar one [23].
PB	Sum of Distances among Page based Fixations	Calculates the total distances among all points.	Fixation distance is longer for unfamiliar user than familiar one [23].
PB	Mean of Path Angles among Page based Fixations	Calculates the average angle that takes place between sequential points according to horizontal axis.	Familiar user look at bigger angles than unfamiliar one [6].
PB	Sum of Path Angles among Page based Fixations	Calculates the sum of angles that take place between sequential points according to horizontal axis.	Familiar user look at bigger angles than unfamiliar one [6].
PB	Page based Fixation Counts per Sequence- based Fixation Counts	Calculates the rate of search task based fixation counts over total fixation counts.	Familiar user have bigger proportion of sequence- based fixation counts over all fixations than unfamiliar one [19].

Table 2.3: Feature-Level Related Work and Gender-Related Hypothesis

Type	Feature	How to Compute	Gender-Related Hypothesis
SB	Scanpath	Shows the sequence of AOIs that a user looks one by one.	There are differences in scanpaths of female and male.
SB	Mean of Sequence based Fixation Durations	Mean of fixation durations over AOIs.	Female's fixation duration is longer than male [7, 21, 3].
SB	Sum of Sequence based Fixation Durations	Sum of fixation durations over AOIs.	Female's fixation duration is longer than male [7, 21, 3].
SB	Sequence based Fixation Counts	Number of fixations over AOIs.	Male make fewer fixations than females [7, 19].
SB	First Fixated AOI	AOI that the participant looks at first.	First Fixated AOI differs in terms of gender.
SB	Percentage of First Fixated AOI	Percentage of the first fixated duration within whole duration.	Duration of First Fixated AOI is different than other fixations as a percentage in terms of gender.
SB	Duration of First Fixated AOI	Duration when the participant looks at the first AOI.	Female's fixation duration on First Fixated AOI is longer than male [7, 3].
PB	Mean of Page based Fixation Durations	Mean of all fixation durations.	Female's fixation duration is longer than male [7, 21, 3].
PB	Sum of Page based Fixation Durations	Sum of all fixation durations.	Female's fixation duration is longer than male [7, 21, 3].
PB	Page based Fixation Counts	Count all fixations	Male make fewer fixations than females [7, 19, 24].
PB	Number of Viewed AoIs per Page based Fixations	Fragments the number of AoIs that the participant views in over number of page based fixations.	There are differences between familiar and unfamiliar user [21, 22].
PB	Mean of Distances between Page based Fixations	Calculates the average of distances among all points.	Fixation distance differs in terms of gender.
PB	Sum of Distances between Page based Fixations	Calculates the total distances among all points.	Fixation distance differs in terms of gender.
PB	Mean of Path Angles between Page based Fixations	Calculates the average angle that takes place between sequential points according to horizantal axis.	Path angle differs in terms of gender.
PB	Sum of Path Angles between Page based Fixations	Calculates the sum of angles that take place between sequential points according to horizantal axis.	Path angle differs in terms of gender.
PB	Page based Fixation Counts per Sequence-based Fixation Counts	Calculates the rate of search task based fixation counts over total fixation counts.	Gender affects the proportion of looking AOI or non-AOI.

Support Vector Machines which was invented by Vapnik in 1982 is simply a hyperplane which is between positive and negative instances with the maximum margin [27]. In addition to SVM, in 1998, Platt came up with a boosted version of SVM; Sequential Minimal Optimization (SMO). Platt had complained from a low working speed of SVM which solves a series of quadratic programming (QP) problems. SMO also solves QP problems by decomposing them into sub-problems and at every step by optimizing values for multipliers [28]. Sequential Minimal Optimization looks more appropriate to classify plenty of independent variables to infer a binary dependent variable.

In addition, the next section will indicate where eye-tracking features are retrieved for both familiarity and gender factors.

2.4 Summary

In this chapter, related studies are provided under three sections; eye tracking, data mining techniques, and gaze features. They compose a tripod on which the thesis stands on. They are examined in detail and help us to see the gap which the thesis intends to fill in and to determine the right gaze features and data mining techniques to train them. The next section will explain how the datasets are extracted from raw data to prepare them for training.

CHAPTER 3

FEATURE EXTRACTION TOOL

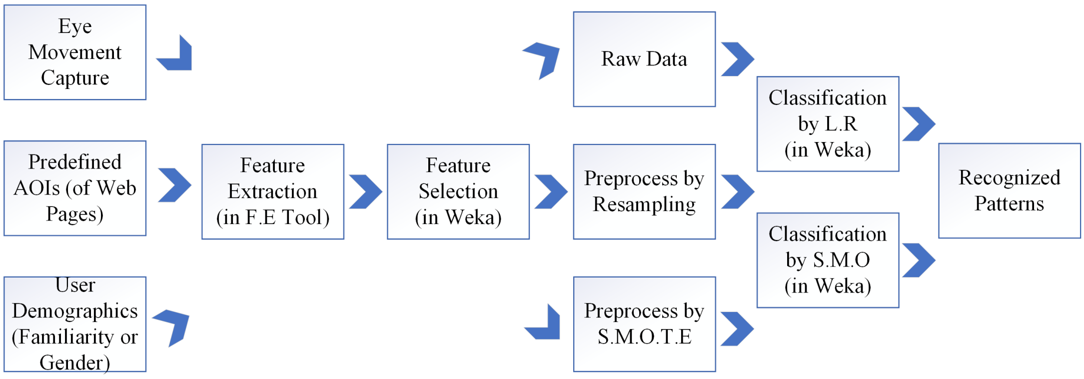
Literature review enabled us to identify eye tracking features (*see Tables* 2.2 and 2.3) that have been used in the related work. Before the classification, their prediction weights for familiarity and gender will be measured separately. In the beginning, eye tracking features need to be extracted from eye tracker output of each user, predefined Areas of Interest (AOI) of each web page and users' demographic file. Since there is no open source tool to do the extraction, it is decided to develop a Feature Extraction Tool in Java environment and present it as an open source in which other researchers may use and modify it for their specific purposes. This chapter defines the Feature Extraction Tool, describes its architecture and implementation in detail.

3.1 Feature Extraction

Feature extraction (FE) is a process analyzing huge volumes of data and extracting the dimensions with excluding repeated factors [29]. He et al. stated that feature extraction means the transformation of original data by keeping the most discriminatory information which improves classification performance [30]. Feature extraction is frequently conducted for pattern recognition and image processing problems. This study attempts to characterize users based on their eye movements over web pages which can be formulated as a pattern recognition problem (*see Figure* 3.1). Thus, it is expected that feature extraction could facilitate training of the data models.

Tables 2.2 and 2.3 consist of eye tracking features utilized in the past studies, feature-related hypotheses and "How to Compute" sections. "How to Compute" section describes how to extract each feature from raw eye movement data. The raw eye movement data is required to extract each eye tracking feature while AOI is used to extract only Sequence-based Features. Moreover, supervised features; familiarity and gender are taken from user's demographic file.

Figure 3.1 Process of Eye Gaze Pattern Recognition



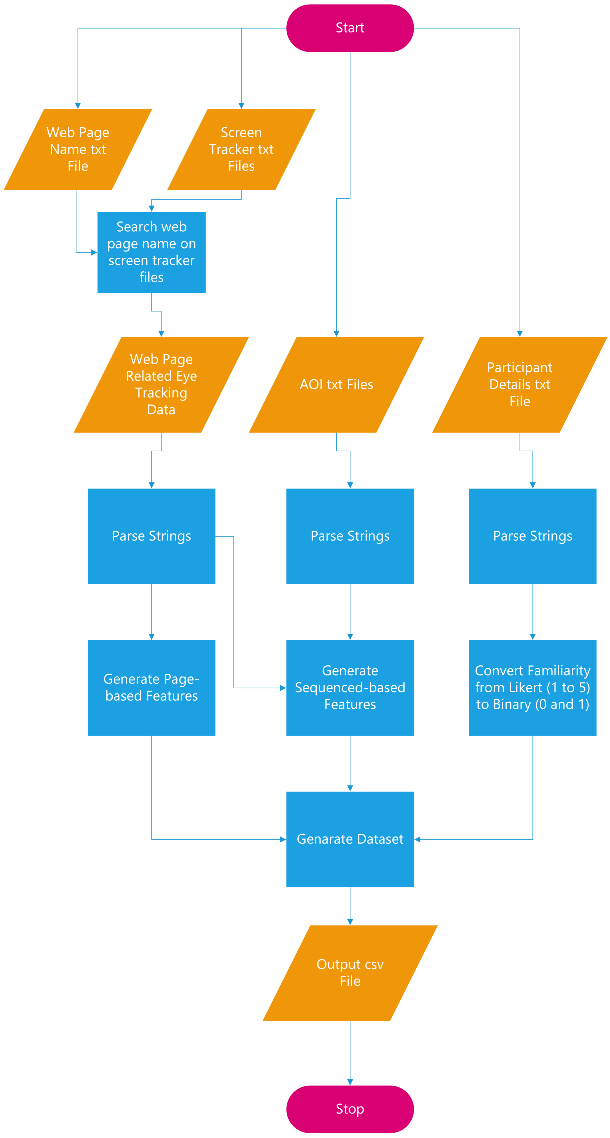
3.2 Architecture

In this section, how Feature Extraction Tool is developed and its components such as inputs and outputs will be discussed. There are three input files (.txt) and an output file (.csv). Firstly, raw eye tracking data records a session of a user as instance-based. Each one consists of six dimensions; FixationIndex, TimeStamp, FixationDuration, MappedFixationPointX, MappedFixationPointY, and StimuliName. Secondly, the AOIs are predefined for each web page. In fact, AOI is determined by the Vision-based Page Segmentation (VIPS) algorithm which uses not only source code but also a visual rendering of web pages [31]. Lastly, user demographic file is processed and used just for supervising factor; either familiarity or gender in this research.

Basically, with respect to computation, there are two types of feature; sequence-based and page-based. While Sequence-based Features need all of three inputs, Page-based Features require just raw eye tracker output and user's demographic file. Although each user has a raw eye tracking data which records in Screen Tracker File folder and each web page has AOIs which records in AOI File folder, there is just one user's demographic file which records in Participant Details folder.

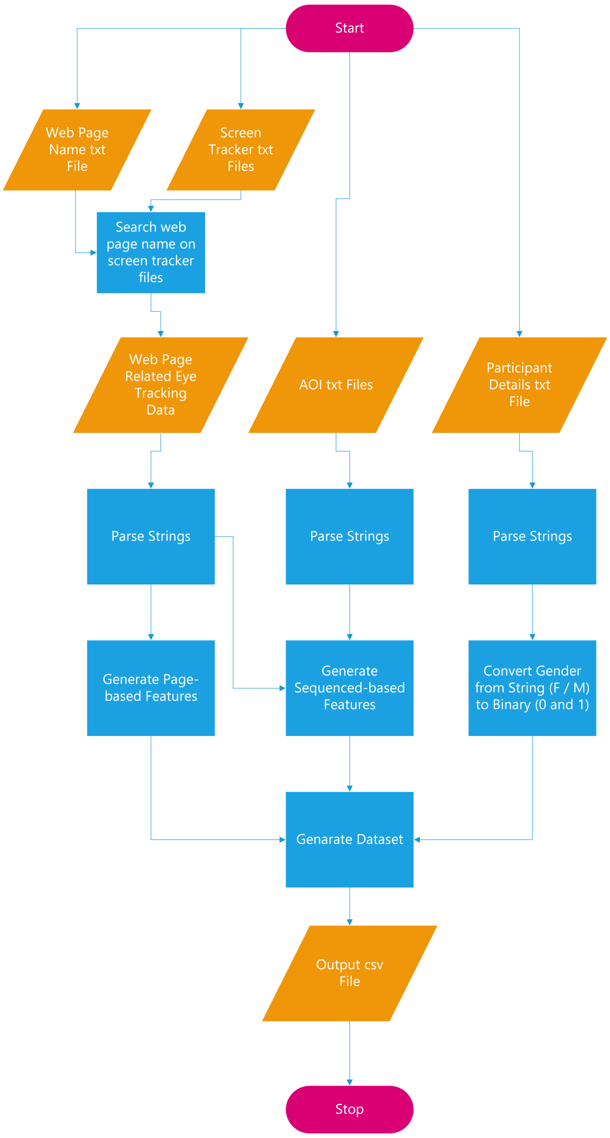
Feature Extraction Tool takes a web page name from 'WebPage.txt ' file that needs to be processed. Then, it seeks down all eye tracker output files of the users according to the 'StimuliName' feature until finding related rows. It takes instances one by one as a string and then parses them based on the spaces between the words. It also takes AOI instances one by one as a string and parses them which includes AOI name, upper left x-y coordinates, and horizontal and vertical side lengths and dedicated letter. Lastly, the tool takes 'ParticipantDetals.txt' file which includes users' ID, gender, age group, educational background and familiarity levels from 1 to 5 for each web page in which 1 represents that user is highly familiar to the web page; but, 5 represents the user has not seen that web page. Moreover, each instance represents a participant in this file. In terms of supervised factors; familiarity and gender, there are two modules of the software. Familiarity module is used to extract users' familiarity from their

Figure 3.2 Feature Extraction Tool FlowChart - Familiarity Module



frequency of use in demographics file (*see Figure 3.2*). If the frequency of use is 1, 2 or 3, in familiarity, it means that the participant uses a web page at least once in a month and the participant is familiar to the web page and encoded to 0. Otherwise, the participant is unfamiliar and encoded to 1. Moreover, gender module is used to take users' gender and encode it to 0 and 1. (*see Figure 3.3*). If the gender needs to be extracted, the one runs and takes gender by enumerating femalea as 0 and males as 1. In the end, the tool generates a dataset by appending all generated features and then export it as an external (.csv) file.

Figure 3.3 Feature Extraction Tool FlowChart - Gender Module



3.3 Implementation and Usage

Feature Extraction Tool is developed with Java. The object-oriented approach is adopted and it may be easily modified to extract new features and supervised factors. However, it is important to give the input files in specific format and extensions. All input files are (.txt) files and output is a (.csv) file. Screen tracker files are in Tobii Pro Studio output format. Actually, Tobii Pro Studio may export lots of eye gaze features; but for this tool, six of them need to be exported which are FixationIndex, TimeStamp, FixationDuration, MappedFixationPointX, MappedFixationPointY, and StimuliName. Tobii Pro Studio exports it as a (.xls) file. Then, it needs to be saved as a (.txt) file (*see Appendix A*). AOIs are manually drawn or automatically extracted by the VIPS algorithm. AOIs are given to this tool as (.txt) files (*see Appendix B*).

Lastly, participant details file is created in MS Excel and converted to a (.txt) file (*see Appendix C*).

After the input files are prepared and placed properly, the web page name, which needs to be extracted, is written into 'WebPage.txt' file and familiarity or gender software module can be executed. The extracted file comprises of 17 columns and rows as much as the number of users.

3.4 Summary

In this chapter, how raw eye movement data is transformed into the datasets for modeling is explained in detail. Feature Extraction Tool is developed to accomplish this transformation and interested researchers may utilize and modify the tool to meet their needs. The next chapter will handle the data modeling phase in which it is explained how to collect and preprocess data, create models, choose algorithms and select features and thus preliminary results before validation is presented.

CHAPTER 4

MODELING STUDY

In the modeling study, by using eye movement data from previous studies, data models are created. The previous studies will be explained in detail in terms of purpose, participants, and procedure. Then, data modeling methodology will be explained to show how to create data models. In order to advance the predictions, feature selection is conducted which will be indicated. Lastly, the modeling results will be presented and discussed with respect to both familiarity and gender.

4.1 Eye Tracking Dataset

For the modeling, an existing eye-tracking dataset is used which was collected from two different eye-tracking studies. The first one aims at extracting scanpath patterns and producing a common scanpath [1]. The second one intends to produce also a common scanpath [2]. Both of them used the same methodology; so, they can be combined and a dataset can be composed. In this dataset, there are 79 participants' data in total. The dataset is publicly available for research purposes.

Table 4.1: Number of Familiar Users for Each Web Page

Web Page	#Familiar	#Unfamiliar
Apple	23	56
BBC	41	38
Yahoo	30	49
AVG	6	73
Babylon	5	74
GoDaddy	1	78

4.1.1 Participants

Both eye-tracking studies were conducted in the following universities; METU NCC and the University of Manchester. Naturally, their participants consist of students mostly; so, 67% of participants are in the age-range 18-24, 25% of them are between

25-34 years old and 8% are older than 35 years old. Similarly, in terms of educational background, they are not evenly distributed into groups because most are students of a bachelor; not graduated. On the other hand, they all are daily web users and with respect to gender, two groups are equal; 51% of them are females when the rests are males. Likewise, in Table 4.1, numbers of familiar and unfamiliar users are presented for six web pages; Apple, AVG, Babylon, BBC, GoDaddy, Yahoo. Familiar and unfamiliar users are determined according to the questionnaire F which asks the participants with how often you visit the web page. On the Likert type scale from 1 to 5, 1 represents daily usage while 5 represents never. It is an assumption that if a participant uses a web page more than once in a month, this means the one is familiar to the web page. Thus, if a participant chooses 1, 2, or 3 for a web page, the one is familiar to the web page, else the one is unfamiliar to the web page.

According to this assumption, the participants are divided into two groups; familiar and unfamiliar users. If the familiar - unfamiliar numbers of users are completely unbalanced, they are canceled because the algorithms cannot classify them properly and cause overfitting in training (*see Table* 4.1). For instance, AVG, Babylon, and GoDaddy are canceled. In this study, familiar and unfamiliar users of Apple, BBC and Yahoo will be taken into account. As a result, in terms of supervised factors, age-range and educational backgrounds of this group cannot be classified in proper. However, gender and familiarity of the users look appropriate for data mining.

4.1.2 Procedure

Six web pages are selected from Alexa Top 100 list and their complexity is measured by the Vicram framework to ensure that they have different complexity levels [32]. Moreover, web pages are divided into AOIs by using the extended version of the Vision-Based Page Segmentation (VIPS) algorithm that automatically defines the AOIs by exploiting web site's source code and visual rendering [31, 33]. Furthermore, Tobii T60 17" built-in eye tracker was used to implement the study and its resolution was 1280 x 1024.

While conducting the eye-tracking study, the procedure has three phases. Firstly, before the implementation, the information sheet is presented to participants which describes the objectives and participants' rights (*see Appendix D*). Then, if they are willing to participate in this study, they need to sign in a consent form (*see Appendix E*). After that, participants should fill in a questionnaire which asks for participant's demographics and how often they visit these web pages (*see Appendix F*). Here, their frequency of use to the web pages is ranked as a Likert type scale (from 1 to 5). While 1 represents daily usage, 5 means that the participant has never seen the web page. This thesis tries to classify users as familiar or unfamiliar as binary. To do so, it is an assumption that if the user chooses 1, 2 or 3, he uses that web page at least once in a month; so, he is a familiar user. On the other hand, he chooses 4 or 5, he is taken as a completely unfamiliar user.

Table 4.2: Searching Task Questions [1, 2]

Web Page	Related Questions
Apple	a. Can you locate the link that allows watching the TV ads relating to iPad mini? b. Can you locate a link labelled iPad on the main menu?
BBC	a. Can you read the first item of Sport News? b. Can you locate the table that shows market data under the Business title?
Yahoo	a. Can you read the titles of the main headlines which have smaller images? b. Can you read the first item under News title?
AVG	a. Can you locate the link which you can download a free trial of AVG Internet Security 2013? b. Can you locate the link which allows you to download AVG Antivirus Free 2013?
Babylon	a. Can you locate the link that you can download the free version of Babylon? b. Can you find and read the names of other products of Babylon?
GoDaddy	a. Can you find a telephone number for technical support and read it? b. Can you locate a text box where you can search for a new domain?

Secondly, within the implementation, the browsing phase aims at the participants to explore each web page in 30 seconds. The specialist does surely not intervene during browsing. Lastly, within implementation also, the searching phase takes a place in which each participant is asked of two questions for each web page (*see Table* 4.2). The questions are neither difficult nor easy to find the answer over the web page. Answering the questions takes maximum 120 seconds. These tasks aim at differentiating users' eye tracking features to capture their familiarity and gender. In this study, for browsing and searching purposes, different datasets are created because different purposes could influence in eye tracking features.

As a result, in the scope of Eraslan et al. studies [1, 2] the raw eye tracking data is collected from 79 participants for 6 web pages. Participants see each web page twice for both browsing and searching purposes. In this thesis, by exploiting the raw eye movement data, firstly, eye tracking features are extracted by Feature Extraction Tool (*see Chapter* 3). Then, the data models are created by conducting data mining algorithms in Weka 3.8.1 tool [1]. The data models show the results by 10-fold cross validation. Section 4.2 will describe how eye tracking data are trained to construct data models for inferring the familiarity and gender.

4.2 Data Modeling Methodology

Creating data models in Weka 3.8.1 tool requires to preprocess data and apply appropriate algorithms properly. In this section, it is explained which algorithms are utilized and how to choose them. Then, to enhance the results, which preprocessing techniques are used and their justifications will be discussed.

[1] `https://www.cs.waikato.ac.nz/ml/weka/.`

The output of data models is whether a participant is familiar or not and similarly, male or female; that is, the output is binary. Therefore, while training, a binary classifier needs to be utilized. There are lots of binary classifiers to train models such as decision tree, k-nearest neighbors (knn) and two class Bayes. However, similar studies have usually used two classification methods; Support Vector Machine and Logistic Regression for binary classification (*see Chapter* 2). Even when Rello et al. tries to detect if a user is dyslexic or not, they used the SVM classifier [8]. Similarly, Yaneva et al. utilizes Logistic Regression to classify users as Autistic or not [9]. Therefore, it is decided to use these algorithms to classify familiarity and gender characteristics of the users. Since the datasets have been trained in Weka 3.8.1 tool, Sequential Minimal Optimization (SMO) is used as it is the fastest way of applying SVM [28].

In the scope of the modeling study, lets shortly look at how the chosen algorithms work and how to train data. As mentioned above, two binary classifiers are applied in this study; SMO and Logistic Regression. Firstly, SMO was discovered by John Platt in 1998 [28]. In this training approach, the algorithm tries to draw a hyperplane between two groups which divides the groups where maximize the distance to the nearest instance from both sides. In fact, it requires to solve a large quadratic programming (QP) optimization problem. Platt says that unlike SVM, SMO breaks this problem into a series of possible QP problems and shortens training time [28]. Secondly, Logistic Regression is conducted in Weka 3.8.1 tool with a ridge estimator [34] which attempts to optimize an m*(k-1) matrix in which k is the number of classes with m attributes.

Before the classification, it is required to preprocess data because preprocessing makes the data smooth and improves the results. However, both raw and preprocessed results will be presented. First of all, in order to apply binary classifiers, familiarity and gender characteristics cannot be a numerical variable. Thus, in Weka 3.8.1 tool, the NumerictoNominal filter is utilized to prepare the dataset for classifying. Furthermore, in the datasets, there are just 79 instances which are a small set to train; so, it is decided to smooth data by Resampling technique. Resampling in Weka 3.8.1 tool means to construct subsample of a dataset with or without replacement [35]. In this study, Resampling is applied with replacement and do not disturb the uniformity of the dataset; for instance, familiar and unfamiliar numbers of instances do not change. Therefore, resampling gets the instances closer and eliminates the outliers.

Lastly, because of the small datasets problem also, Synthetic Minority Oversampling Technique (SMOTE) is applied to datasets. SMOTE was discovered by Chawla et. al in 2002 to solve class imbalance problems [36]. In this technique, minority class is over-sampled by taking each minority class instances into account to create synthetic examples joining any/all k minority class nearest neighbors. In this study, by using SMOTE with 5-nearest neighbors, it creates synthetic instances until 50% of minority class instances.

With these data mining methods and preprocessing techniques, the data models are created for each web page. In Section 4.4, the results will be presented with respect to the web page, algorithm, purpose, preprocessing, and classification factors. Before presentation of the results, the feature selection which tries to eliminate weak features and enhance the results will be presented.

4.3 Feature Selection

Table 4.3: Feature Selection by Information Gain

	Familiarity Factor		Gender Factor	
Attributes	**Average Merit**	**Tolerance**	**Average Merit**	**Tolerance**
Scanpath	0.94	0.003	0.96	0.012
Mean of Sequence based Fixation Durations	0.94	0.003	0.97	0.011
Sum of Sequence based Fixation Durations	0.93	0.006	0.96	0.012
Mean of Page based Fixation Durations	0.94	0.003	0.97	0.011
Sum of Page based Fixation Durations	0.94	0.003	0.96	0.012
Sequence based Fixation Counts	0.43	0.027	0.45	0.035
Page based Fixation Counts	0.52	0.028	0.60	0.026
Number of Viewed AoIs per Page Based Fixations	0.81	0.023	0.85	0.22
First Fixated AoI	0.09	0.017	0.09	0.018
Percentage of First Fixated AoI	0.93	0.006	0.96	0.012
Duration of First Fixated AoI	0.37	0.029	0.41	0.031
Mean of Distances between Page based Fixations	0.94	0.003	0.97	0.011
Sum of Distances between Page based Fixations	0.94	0.003	0.97	0.011
Mean of Path Angles between Page based Fixations	0.94	0.003	0.97	0.011
Sum of Path Angles between Page based Fixations	0.94	0.003	0.97	0.011
Page based Fixation Counts per Task based Fixation Counts	0.88	0.017	0.93	0.019

In this study, we investigated 16 features that were identified in the literature to predict the familiarity or gender characteristics (*see Tables* 2.2 and 2.3). However, all of the attributes cannot give benefit equally for training. There are some ways to determine the attributes' estimation power. We conduct Information Gain method to determine the predictivity of attributes. Table 4.3 shows Information Gain's results. In this table, average merit shows the value (0-1) which indicates the percentage of the relationship between the attribute and supervised factor; familiarity or gender. According to Demisse, in average merit, 0.5 is a relevancy threshold [37]. In detail, average merits under 0.5 are irrelevant to target feature; so, in this study, I subtract tolerance value from average merit and if the result is under 0.5 threshold, this attribute is canceled in the dataset. As a result, I cancel the four attributes for the familiarity factor when three of them are not used for gender factor. Canceled attributes are highlighted by red color in Table 4.3.

In order to prove power of training with selected features, Table 4.4 shows both of

Table 4.4: Proof of Feature Selection Power

		with All Features			with Selected Features		
	Algorithms	**Apple**	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	68.35%	43.04%	53.16%	70.89%	55.69%	62.02%
	Precision	0.497	0.431	0.454	0.709	0.564	0.620
	Recall	0.684	0.430	0.532	0.709	0.557	0.620
	F-measure	0.576	0.431	0.474	0.830	0.525	0.766
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	88.61%	78.48%	86.08%	88.61%	72.15%	88.60%
	Precision	0.892	0.785	0.860	0.902	0.722	0.904
	Recall	0.886	0.785	0.861	0.886	0.722	0.886
	F-measure	0.879	0.785	0.860	0.876	0.722	0.881
	Number of Instances	79	79	79	79	79	79
SMOTE	**Accuracy**	74.44%	62.24%	68.09%	75.55%	71.43%	69.14%
	Precision	0.763	0.613	0.687	0.825	0.728	0.806
	Recall	0.744	0.622	0.681	0.756	0.714	0.691
	F-measure	0.719	0.602	0.676	0.717	0.716	0.653
	Number of Instances	90	98	94	90	98	94

results trained by Sequential Minimal Optimization with browsing data for familiarity factor. It is seen that feature selection boosts accuracy values up to 3.5% as average. Therefore, in Section 4.4, the results will be exhibited for both familiarity and gender characteristics with the selected features.

Table 4.5: Familiarity Models by Logistic Regression

		Browsing			Searching		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	67.09%	54.43%	51.89%	73.41%	49.36%	60.75%
	Precision	0.585	0.545	0.318	0.709	0.470	0.559
	Recall	0.671	0.544	0.519	0.734	0.494	0.608
	F-measure	0.604	0.521	0.424	0.696	0.434	0.537
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	84.81%	74.68%	87.34%	78.48%	73.41%	88.60%
	Precision	0.846	0.748	0.885	0.785	0.737	0.904
	Recall	0.848	0.747	0.873	0.785	0.734	0.886
	F-measure	0.841	0.746	0.868	0.785	0.732	0.881
	Number of Instances	79	79	79	79	79	79
SMOTE	**Accuracy**	86.67%	75.51%	76.59%	88.88%	68.36%	78.72%
	Precision	0.869	0.760	0.804	0.888	0.680	0.799
	Recall	0.867	0.755	0.734	0.889	0.684	0.787
	F-measure	0.867	0.756	0.714	0.888	0.678	0.684
	Number of Instances	90	98	94	90	98	94

4.4 Results

The results show the performance of the modeling study which indicates preliminary results and they are very promising. The results are presented in six tables; for

Table 4.6: Familiarity Models by Sequential Minimal Optimization

		Browsing			**Searching**		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	70.89%	55.69%	62.02%	70.88%	45.56%	60.75%
	Precision	0.709	0.564	0.620	0.709	0.400	0.382
	Recall	0.709	0.557	0.620	0.709	0.456	0.608
	F-measure	0.830	0.525	0.766	0.830	0.386	0.469
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	88.61%	72.15%	88.60%	79.74%	70.88%	88.60%
	Precision	0.902	0.722	0.904	0.799	0.709	0.904
	Recall	0.886	0.722	0.886	0.797	0.709	0.886
	F-measure	0.876	0.722	0.881	0.798	0.709	0.881
	Number of Instances	79	79	79	79	79	79
SMOTE	**Accuracy**	75.55%	71.43%	69.14%	74.44%	64.28%	68.08%
	Precision	0.825	0.728	0.806	0.819	0.642	0.776
	Recall	0.756	0.714	0.691	0.744	0.643	0.681
	F-measure	0.717	0.716	0.653	0.701	0.642	0.644
	Number of Instances	90	98	94	90	98	94

familiarity and gender models by Logistic Regression and Sequential Minimal Optimization.

According to Vicram framework [32], Apple's page complexity level is low although Yahoo's is medium and BBC's is high. Different complexity levels help us to assess how the web page's visual complexity affects user's familiarity and gender. Furthermore, browsing and searching results which enable to evaluate the difference between searching and browsing in terms of familiarity and gender factors are presented separately. In addition, as explained in Section 4.2, the results of raw data is presented while resampled and synthesized by SMOTE data are exhibited. Thanks to those preprocessing, it is seen that the data models are strengthened.

4.4.1 Familiarity Data Models

Tables 4.5 and 4.6 show the results of the modeling for familiarity factor. Because of the imbalanced number of familiar and unfamiliar users to three web pages; AVG, Babylon and GoDaddy, data models are overfitted and all of their accuracy values are almost 100%. Then, they are eliminated and are not shown in the result tables. The rest data models show that both SMO and Logistic Regression algorithms work well because their average of all values is higher than 50% threshold which means their predictivity is good.

In the raw data models, searching and browsing values look similar and likely, Logistic Regression and SMO algorithms produce similar results. In terms of visual complexity, there is no consistent evidence to claim that visual complexity is an influencer while predicting familiarity. The lowest accuracy in the raw data models is 45.56% which belongs to BBC searching data trained by SMO. The highest accuracy value is 73.41% which belongs to Apple searching data trained by Logistic Regression. Precision and recall measures are balanced for all raw data models

In the resampled data models, browsing results are 3% better than searching ones as average. Logistic Regression and SMO algorithms produce similar results. Moreover, there is no clear pattern to claim that visual complexity influences the prediction of familiarity with resampled data. The lowest accuracy in the raw data models is 70.88% which belongs to BBC searching data trained by SMO. The highest accuracy value is 88.60% which belongs to Yahoo browsing and searching data trained by both algorithms and Apple browsing data trained by SMO. Precision and recall measures are balanced for all resampled data models.

In the synthetic oversampled data models, browsing results are 2% better than searching results as average. Logistic Regression produces almost 9% better results than the SMO algorithm as average. The best accuracy values are produced on the pages with the lowest complex in the oversampled data models. The lowest accuracy in the raw data models is 64.28% which belongs to BBC searching data trained by SMO. The highest accuracy value is 88.88% which belongs to Apple searching data trained by Logistic Regression. Precision and recall measures are balanced for all synthetic oversampled data models.

Table 4.7: Gender Models by Logistic Regression with Browsing Data

		Browsing					
		Apple	**AVG**	**Babylon**	**BBC**	**GoDaddy**	**Yahoo**
	Visual Complexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	39.24%	44.30%	48.10%	43.03%	49.36 %	58.22%
	Precision	0.390	0.441	0.479	0.431	0.493	0.582
	Recall	0.392	0.443	0.481	0.430	0.494	0.582
	F-measure	0.390	0.439	0.476	0.430	0.49	0.581
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	72.15%	69.62%	72.15%	82.27%	77.21%	70.88%
	Precision	0.726	0.696	0.722	0.826	0.775	0.715
	Recall	0.722	0.696	0.722	0.823	0.772	0.709
	F-measure	0.720	0.696	0.722	0.822	0.772	0.706
	Number of Instances	79	79	79	79	79	79
SMOTE	**Accuracy**	69.38%	71.42%	72.44%	72.44%	77.55%	79.59%
	Precision	0.694	0.714	0.722	0.732	0.776	0.796
	Recall	0.694	0.714	0.724	0.724	0.776	0.796
	F-measure	0.694	0.714	0.723	0.726	0.776	0.796
	Number of Instances	98	98	98	98	98	98

4.4.2 Gender Data Models

Tables 4.7, 4.8, 4.9, and 4.10 show the results of the gender data models which are created for modeling. Gender data models are produced for all six web pages. Similar to the familiarity factor, preprocessed data models' results are seen better than raw

Table 4.8: Gender Models by SMO with Browsing Data

		Browing					
		Apple	**AVG**	**Babylon**	**BBC**	**GoDaddy**	**Yahoo**
	Visual Complexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	37.97%	48.10%	49.36%	58.16%	46.83%	56.96%
	Precision	0.378	0.475	0.492	0.566	0.468	0.581
	Recall	0.380	0.481	0.494	0.582	0.468	0.570
	F-measure	0.379	0.460	0.488	0.566	0.468	0.550
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	68.35%	65.82%	67.08%	79.74%	72.15%	69.62%
	Precision	0.693	0.663	0.671	0.804	0.722	0.709
	Recall	0.684	0.658	0.671	0.797	0.722	0.696
	F-measure	0.679	0.655	0.671	0.796	0.722	0.691
	Number of Instances	79	79	79	79	79	79
SMOTE	**Accuracy**	64.28%	59.18%	61.22%	59.18%	70.40%	67.34%
	Precision	0.639	0.578	0.604	0.592	0.701	0.680
	Recall	0.643	0.592	0.612	0.592	0.704	0.673
	F-measure	0.640	0.579	0.606	0.592	0.702	0.675
	Number of Instances	98	98	98	98	98	98

Table 4.9: Gender Models by Logistic Regression with Searching Data

		Searching					
		Apple	**AVG**	**Babylon**	**BBC**	**GoDaddy**	**Yahoo**
	Visual Complexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	45.56%	35.44%	59.49%	54.43%	52.56%	63.29%
	Precision	0.454	0.355	0.610	0.544	0.524	0.661
	Recall	0.456	0.354	0.595	0.544	0.526	0.633
	F-measure	0.453	0.354	0.583	0.544	0.523	0.614
	Number of Instances	79	79	79	79	79	79
Resampling	**Accuracy**	70.88%	69.62%	81.01%	75.94%	82.05%	74.68%
	Precision	0.710	0.698	0.82	0.760	0.828	0.748
	Recall	0.709	0.696	0.81	0.759	0.821	0.747
	F-measure	0.708	0.695	0.808	0.759	0.819	0.747
	Number of Instances	79	79	79	79	78	79
SMOTE	**Accuracy**	68.36%	66.32%	76.53%	77.55%	75.25%	81.63%
	Precision	0.688	0.675	0.773	0.785	0.751	0.844
	Recall	0.684	0.663	0.765	0.776	0.753	0.816
	F-measure	0.685	0.666	0.767	0.777	0.752	0.818
	Number of Instances	98	98	98	98	97	98

Table 4.10: Gender Models by SMO with Searching Data

		Searching					
		Apple	**AVG**	**Babylon**	**BBC**	**GoDaddy**	**Yahoo**
	Visual Complexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	43.03%	44.30%	58.22%	50.63%	46.15%	59.49%
	Precision	0.425	0.418	0.582	0.505	0.450	0.625
	Recall	0.430	0.443	0.582	0.506	0.462	0.595
	F-measure	0.422	0.403	0.582	0.487	0.441	0.566
	Number of Instances	79	79	79	79	78	79
Resampling	**Accuracy**	69.62%	68.35%	78.48%	68.35%	74.35%	73.41%
	Precision	0.697	0.689	0.800	0.698	0.753	0.738
	Recall	0.696	0.684	0.785	0.684	0.744	0.734
	F-measure	0.696	0.681	0.782	0.677	0.74	0.733
	Number of Instances	79	79	79	79	78	79
SMOTE	**Accuracy**	58.16%	62.24%	65.30%	69.38%	67.01%	72.44%
	Precision	0.566	0.624	0.649	0.692	0.680	0.728
	Recall	0.582	0.622	0.653	0.694	0.670	0.724
	F-measure	0.566	0.623	0.65	0.693	0.672	0.726
	Number of Instances	98	98	98	98	97	98

data models. However, generally, the values are worse than familiarity.

In the raw data models, searching values are better than browsing and both algorithms produce similar results. In terms of visual complexity, there is no consistent evidence to claim that visual complexity is an influencer while predicting gender. The lowest accuracy in the raw data models is 35.44% which belongs to AVG searching data trained by Logistic Regression. The highest accuracy value is 63.29% which belongs to Yahoo searching data trained by Logistic Regression. Precision and recall measures are balanced for all raw data models.

In the resampled data models, searching values are also better than browsing values. Logistic regression results are 4% better than SMO as average. In terms of visual complexity, the results do not show a pattern. The lowest accuracy value in the resampled data models is 65.82% which belongs to AVG browsing data trained by SMO. The highest accuracy value is 82.27% belongs to BBC browsing data trained by Logistic Regression. Precision and recall measures are balanced for all resampled data models.

In the synthetic oversampled data models, searching and browsing values are almost equal to each other. Logistic regression results are 8% better than SMO as average. In terms of visual complexity, the higher complex web page has a better result than the lower complex one. The lowest accuracy value in the oversampled data models is 58.16% which belongs to Apple searching data trained by SMO. The highest accuracy value is 81.63% belongs to Yahoo searching data trained by Logistic Regression. Precision and recall measures are balanced for all oversampled data models.

4.5 Summary and Conclusion

In this chapter, how to, what, and why questions are tried to ask and answer about modeling study. Section 4.1 explains the existing eye-tracking studies which were conducted by Eraslan et al. for different research purposes [1, 2]; but, the collected data was used to conduct modeling and extract data models. There are totally 79 participants' eye movement data from two eye-tracking studies. In the two studies, the same procedure was followed in order to maintain the unity and integrity of two datasets.

Section 4.2 presents the modeling methodology. Based on the previous studies' data and algorithms, Logistic Regression and Sequential Minimal Optimization are decided to be utilized to construct data models. In order to boost the results by a reliable method, feature selection technique, Information Gain, is conducted and the study eliminates weak attributes while predicting gender and familiarity factors in Section 4.3.

In Section 4.4, familiarity and gender data models are exhibited. The models' details and inferences are discussed. In overall, the results look promising and valuable to be validated by another eye-tracking study. A validation study is conducted to see if the results are coincidence or not. A new eye-tracking study is conducted with 20 participants and the same procedure of the previous two studies. This is to ensure that we have data integrity. Next chapter 5 will present to validation study in detail.

CHAPTER 5

VALIDATION STUDY

The modeling results are seen as promising and support the research questions; however, it requires to prove its repeatability because artificial intelligence systems are alive; in other words, they collect and analyze data progressively. Therefore, an eye tracking study is conducted again and the procedure of the previous two studies is followed. In this study, 20 users participate in the study. Eye tracking study will be described and in terms of similarity and differences with the previous eye tracking studies will also be discussed in detail. Then, the last study's data (20 participants) are added to the previous data (79 participants) to create the new data models and the existing data models try to predict new dataset's familiarity and gender characteristics (*see Section* 5.3). According to the previous feature selection, weak features are canceled from datasets (*see Table 4.3*).

5.1 Eye Tracking Study

In order to validate the modeling results and prove their consistency, an eye tracking study is conducted again. Although the procedure is followed as the same, some conditions differ from the previous two studies. In this section, how the eye tracking study is conducted will be explained in step by step.

Table 5.1: Familiars and Unfamiliars to Web Pages

Web Page	**#Familiar**	**#Unfamiliar**
Apple	11	9
BBC	11	9
Yahoo	12	8
AVG	6	14
Babylon	3	17
GoDaddy	1	19

5.1.1 Participants

An eye tracking study is conducted in Ankara from October 2018 to November 2018. Participants is found by opportunity sampling technique which means that asking available members of the population if they join in the research or not [38]. Fortunately, all users are daily web users. In detail, 8 of the participants are females when the rests are males. If three age-groups are enumerated, 18-24, 25-34, and 35-54 will become group1, group2, and group3, respectively. Group1 has 2 users, group2 has 10 users, and group3 has 8 users. Furthermore, 40% of them are graduated from high school while 40% also took undergraduate degree. In the rest of them, 15% have a graduate degree and 5% just completed middle school. Appendix N shows 20 participant's demographics information as a table.

Table 5.2: Participants with Calibration Problem

	Participant ID	Number of Participants with Problem	Number of Participants without Problem
Apple - Browsing	P9, P14, P17, P18, P20	5	15
Apple - Searching	P4, P11, P14, P16	4	16
AVG - Browsing	P9, P14, P16, P17	4	16
AVG - Searching	-	0	20
Babylon - Browsing	-	0	20
Babylon - Searching	-	0	20
BBC - Browsing	P9, P10, P14	3	17
BBC - Searching	P2, P6, P7, P9, P11 P14, P15, P16, P19, P20	10	10
GoDaddy - Browsing	-	0	20
GoDaddy - Searching	P2, P9	2	18
Yahoo - Browsing	P4, P14	2	18
Yahoo - Searching	P14	1	19

Table 5.1 shows participants' familiarity to the web pages used in the study. Similar to the previous studies, in this study, AVG, Babylon and GoDaddy pages' familiar - unfamiliar users are unbalanced. They are also eliminated for familiarity. Thus, while gender data models are validated for all 6 web pages, familiarity data models are validated for just 3 web pages. In addition, because of eye tracker calibration problems, some sessions are not recorded properly. Table 5.2 shows which participants have the problem with the mobile eye tracker in which web pages.

5.1.2 Procedure

Similar to the previous two studies, the procedure has got three phases; introduction phase, browsing phase, and searching phase. In the introduction phase, sample user is informed with an information sheet in Turkish (*see Appendix G*). Information sheet includes both purpose of the study and participants rights. If one would like to join in the study, he should sign the consent form which is prepared also in Turkish (*see Appendix H*). Lastly, a questionnaire is provided to the users in Turkish (*see Appendix I*) which collects participant's gender, age, graduation and especially frequency of the web page use. In order to keep it consistent, frequency of use options is not converted to binary; but the Likert Type scale is used for the frequency of use. Then familiarity is inferred from frequency of use. Although three web pages are canceled for familiarity modeling because of unequal numbers of familiar - unfamiliar users, canceled web pages are also presented in this eye tracking study because integrity and consistency between previous two studies and this study need to be maintained.

Table 5.3: Searching Task Questions in Turkish

Web Page	**Related Questions**
Apple	a. iPad mini ile ilgili TV reklamını izlemek için tıklar mısınız? b. Ana menüde iPad linkini bulup tıklar mısınız?
BBC	a. Spor haberlerinden ilk olanı okur musunuz? b. Business başlığı altında piyasa verilerini içeren tabloyu bulur musunuz?
Yahoo	a.Yanında küçük simgeler bulunan ana başlıkları okur musunuz? b. Haberler başlığı altındaki ilk haberi okur musunuz?
AVG	a. AVG Internet Security 2013'ün free deneme sürümünü indirmek üzere tıklar mısınız? b.AVG Antivirüs Free 2013'ü indirmek üzere tıklar mısınız?
Babylon	a. Babylon'un free versiyonunu indirmek üzere tıklar mısınız? b. Babylon'a ait diğer ürünlerin isimlerini bulup okur musunuz?
GoDaddy	a. Teknik desteğe ait telefon numarasını bulup okur musunuz? b. Yeni bir alan adı aramak için yazar mısınız?

Then, within implementation, each participant is browsing each web page for 30 seconds. In browsing phase, the participants are not intervened to prevent confusion. Afterthat, in searching phase, two questions are asked about each web page in Turkish (*see Table* 5.3). For each web page, the duration was no longer than 120 seconds.

5.1.3 Equipment

Although the procedure and materials used are the same with the previous studies, there is an important difference in terms of equipment. This study could not be conducted in a laboratory environment. In the previous studies, Tobii T60 17" built-in eye tracker was used to implement the study and its resolution was 1280 x 1024 although in this study, Tobii X2-60 mobile eye tracker is connected to a personal laptop which is Dell Latitude 7280 notebook with 12.5" display and 1366 x 768 resolution. All participants use this laptop and this eye tracker. In order to record sessions and take outputs, Tobii Pro Studio 3.4.8.1348 is installed and utilized. While extracting eye movement data from Tobii Pro Studio, I-VT filter with default settings is used which means that there is no limit to eliminate any fixations from the dataset. In the previous two studies, the outputs were similarly taken by I-VT filter of Tobii Pro Studio with the default settings.

Unfortunately, because of the mobility of the eye tracker used, sometimes calibration problems were experienced when we analyzed the data, we identified some calibration problems.

5.1.4 Materials

Information sheet is different from the previous studies' in terms of both language and content (*see Appendix* G). It is prepared in Turkish although the originals are in English. Moreover, the purpose is different from the previous two studies. Consent form and questionnaire are the same with the previous ones in terms of content (*see Appendix* H, I); but, their language is in Turkish differently. Furthermore, searching questions are prepared in Turkish (*see Table* 5.3); however, their meanings are the same with the previous ones.

Table 5.4: Statistical Analysis of Familiarity

Features	Familiarity related Hypothesis (H1)	Browsing			Searching		
		Apple (t or w)	BBC (t or w)	Yahoo (t or w)	Apple (t or w)	BBC (t or w)	Yahoo (t or w)
Mean of Sequence based Fixation Durations	Familiar user's mean of sequence based fixation durations is longer than unfamiliar user's.	w: 31 d: 0.74 df: NA	w: 23 d: 0.3 df: NA	t: 0.32 d: 0.16 df: 12	w: 32 d: 0.42 df: NA	t: 1.15 d: 0.68 df: 4	t: -0.5 d: -0.26 df: 9
Sum of Sequence based Fixation Durations	Familiar user's sum of sequence based fixation durations is longer than unfamiliar user's.	w:22 d: 0.05 df: NA	t: -1.12 d: -0.59 df: 9	t: 1.23 d: 0.61 df: 13	t: -0.55 d: -0.28 df: 12	t: -1.39* d: -1.01 df: 4	t: 0.32 d: 0.21 df: 5
Sequence based Fixation Counts	Familiar user makes fewer fixations than unfamiliar user.	t: -0.18 d: -0.1 df: 10	t: -0.57 d: -0.3 df: 7	t: 1.18 d: 0.58 df: 13	t: -1.35 d: -0.69 df: 12	t: -2.00* d: -1.51 df: 3	t: 0.34 d: 0.23 df: 5
Percentage of First Fixated AOI	Duration of First Fixated AOI is different significantly than other fixations as a percentage in terms of familiarity.	w: 23 d: 0.09 df: NA	w: 23 d: -0.06 df: NA	w: 37 d: -0.007 df: NA	w: 26 d: 0.22 df: NA	t: 0.8 d: 0.56 df: 5	t: -0.39 d: -0.19 df: 10
Duration of First Fixated AOI	Familiar user's fixation duration on First Fixated AOI is longer than unfamiliar user's.	w: 27 d: 0.59 df: NA	w: 22 d: -0.19 df: NA	w: 39 d: -0.006 df: NA	w: 20 d: -0.30 df: NA	t: -0.74 d: -0.55 df: 3	t: -1.53* d: -0.85 df: 8
Mean of Page based Fixation Durations	Familiar user's fixation duration is longer than unfamiliar user's.	t: 0.28 d: 0.15 df: 8	w: 29 d: 0.45 df: NA	t: -1.09 d: -0.58 df: 9	w: 27 d: 0.34 df: NA	w: 12 d: 0.62 df: NA	w: 22 d: -0.81 df: NA
Sum of Page based Fixation Durations	Familiar user's fixation duration is longer than unfamiliar user's.	w: 24 d: 0.41 df: NA	w: 15 d: -0.78 df: NA	t: 0.36 d: 0.18 df: 12	t: -0.71 d: -0.36 df: 12	t: -2.24* d: -1.67 df: 3	t: -0.92 d: -0.45 df: 10
Page based Fixation Counts	Familiar user makes fewer fixations than unfamiliar user.	t: 0.55 d: 0.29 df: 9	w: 25 d: -0.42 df: NA	t: 1.45 d: 0.68 df: 13	t: -1.20 d: -0.61 df: 12	t: -2.44* d: -1.83 df: 3	t: -0.09 d: -0.05 df: 9
Number of Viewed AOIs per Page based Fixations	There is a significant difference in number of viewed AOIs per page based fixation counts of familiar and unfamiliar users.	t: 0.15 d: 0.09 df: 7	t: 0.41 d: 0.21 df: 12	t: 0.40 d: 0.21 df: 10	t: 0.11 d: 0.05 df: 10	t: 4.07* d: 2.50 df: 5	t: 1.73 d: 0.68 df: 13
Mean of Distances among Page based Fixations	Unfamiliar user's mean of fixation distances are longer than familiar user's.	t: -1.17 d: -0.66 df: 9	t: -0.15 d: -0.07 df: 12	w: 30 d: 0.18 df: NA	w: 23* d: -0.92 df: NA	t: -1.17 d: -0.66 df: 9	w: 42 d: 0.14 df:NA
Sum of Distances among Page based Fixations	Unfamiliar user's sum of fixation distances are longer than familiar user's	t: -0.27 d: -0.14 df: 10	t: -0.44 d: -0.23 df: 12	t: -0.99 d: 0.46 df: 12	t: -1.66* d: -0.83 df: 11	t: -1.95* d: -1.46 df: 3	t: -0.20 d: -0.11 df: 7
Mean of Path Angles among Page based Fixations	Familiar user's mean of path angles are bigger than unfamiliar user's.	t: -0.44 d: -0.25 df: 8	t: -1.89* d: -0.97 df: 12	t: 0.41 d: 0.21 df: 12	t: -1.27 d: -0.68 df: 10	t: 0.02 d: 0.01 df: 6	t: 1.17 d: 0.58 df: 11
Sum of Path Angles among Page based Fixations	Familiar user's sum of path angles are bigger than unfamiliar user's.	t: -0.08 d: -0.05 df: 7	t: -1.7* d: -0.83 df: 9	t: 0.12 d: 0.06 df: 13	t: -2.44* d: -1.23 df: 12	t: 0.16 d: 0.12 df: 3	t: 1.36 d: 0.63 df: 13
Page based Fixation Counts per Sequence based Fixation Counts	Familiar user have bigger proportion of sequence- based fixation counts over all fixations than unfamiliar user.	w: 18 d: 0.47 df: NA	t: 0.27 d: 0.14 df: 12	w: 23 d: -0.40 df: NA	w: 39 d: 0.61 df: NA	t: -1.28* d: -0.87 df: 5	w: 17* d: -0.85 df: NA

*p < 0.05

Table 5.5: Means and Standard Deviations of Familiarity

Features	Familiarity related Hypothesis (H1)	Browsing Apple (m, sd)	Browsing BBC (m, sd)	Browsing Yahoo (m, sd)	Searching Apple (m, sd)	Searching BBC (m, sd)	Searching Yahoo (m, sd)
Mean of Sequence based Fixation Durations	Familiar user's mean of sequence based fixation durations is longer than unfamiliar user's.	f: (463, 238) u: (331, 102)	f: (401, 164) u: (365, 53)	f: (348, 53) u: (339, 54)	f: (387, 137) u: (336, 70)	f: (488, 291) u: (337, 28)	f: (343, 56) u: (358, 60)
Sum of Sequence based Fixation Durations	Familiar user's sum of sequence based fixation durations is longer than unfamiliar user's.	f: (5249, 3034) u: (5085, 2539)	f: (17044, 5450) u: (19720, 3416)	f:(15317, 7809) u:(10754, 6950)	f:(12216, 10185) u: (15188, 10418)	f:(18736, 9696) u:(33674, 19580)	f:(18508, 5276) u:(16449, 15020)
Sequence based Fixation Counts	Familiar user makes fewer fixations than unfamiliar user.	f: (14, 8) u: (14,7)	f: (49, 22) u: (54, 9)	f: (45, 24) u: (32, 21)	f: (29, 12) u: (42, 18)	f: (40, 12) u: (102, 60)	f: (57, 13) u: (47, 44)
Percentage of First Fixated AOI	Duration of First Fixated AOI is different significantly than other fixations as a percentage in terms of familiarity.	f: (0.01, 0.009) u: (0.01, 0.008)	f: (0.01, 0.007) u: (0.01, 0.008)	f: (0.01, 0.007) u: (0.01, 0.01)	f: (0.01, 0.01) u: (0.01, 0.007)	f: (0.01, 0.006) u: (0.01, 0.009)	f: (0.009, 0.007) u: (0.01, 0.006)
Duration of First Fixated AOI	Familiar user's fixation duration on First Fixated AOI is longer than unfamiliar user's.	f: (386, 272) u: (269, 82)	f: (326, 207) u: (370, 243)	f: (324, 179) u: (325, 327)	f: (288, 180) u: (362, 288)	f: (303, 138) u: (504, 524)	f: (390, 206) u: (590, 279)
Mean of Page based Fixation Durations	Familiar user's fixation duration is longer than unfamiliar user's.	f: (396, 42) u: (384,101)	f: (451, 283) u: (362, 49)	f: (364, 37) u: (392, 60)	f: (393, 127) u: (360, 43)	f: (424, 170) u: (342, 36)	f: (388, 59) u: (530, 289)
Sum of Page based Fixation Durations	Familiar user's fixation duration is longer than unfamiliar user's.	f: (27622, 2077) u: (25014, 8268)	f: (25177, 3552) u: (27454, 2192)	f: (26841, 2850) u: (26275, 3198)	f: (29222, 14856) u: (34925, 15898)	f: (20727, 7990) u: (49037, 24141)	f: (50623, 25239) u: (62053, 24053)
Page based Fixation Counts	Familiar user makes fewer fixations than unfamiliar user.	f: (70, 11) u: (65, 22)	f: (68, 24) u: (77, 16)	f: (74, 12) u: (67, 7)	f: (73, 31) u: (95, 39)	f: (51, 19) u: (146, 75)	f: (129, 58) u: (132, 62)
Number of Viewed AOIs per Page based Fixations	There is a significant difference in number of viewed AOIs per page based fixation counts of familiar and unfamiliar users.	f: (0.07, 0.04) u: (0.07, 0.02)	f: (0.09, 0.02) u: (0.08, 0.02)	f: (0.08, 0.03) u: (0.07, 0.04)	f: (0.06, 0.02) u: (0.06, 0.01)	f: (0.14, 0.04) u: (0.06, 0.01)	f: (0.07, 0.04) u: (0.04, 0.01)
Mean of Distances among Page based Fixations	Unfamiliar user's mean of fixation distances are longer than familiar user's.	f: (170, 48) u: (198, 36)	f: (187, 47) u: (191, 61)	f: (193, 73) u: (181, 42)	f: (151, 24) u: (183, 40)	f: (170, 48) u: (198, 36)	f: (140, 28) u: (136, 41)
Sum of Distances among Page based Fixations	Unfamiliar user's sum of fixation distances are longer than familiar user's	f: (11991, 4067) u: (12693,5222)	f: (13418, 6184) u: (14784, 5702)	f: (14219, 5457) u: (12088, 3018)	f: (11290, 5600) u: (17882, 9466)	f: (9312, 4271) u: (24910, 15528)	f: (17782, 8309) u: (18890, 11598)
Mean of Path Angles among Page based Fixations	Familiar user's mean of path angles are bigger than unfamiliar user's.	f: (7, 14) u: (10, 10)	f: (-11, 10) u: (-0.07, 11)	f: (-9, 10) u: (-11, 11)	f: (0.3, 11) u: (6, 7)	f: (-0.6, 11) u: (-0.8, 10)	f: (4, 10) u: (-1.5, 9)
Sum of Path Angles among Page based Fixations	Familiar user's sum of path angles are bigger than unfamiliar user's.	f: (463, 931) u: (500, 510)	f: (-571, 395) u: (54, 948)	f: (-754, 852) u: (-808, 816)	f: (-196, 662) u: (847, 978)	f: (-31, 451) u: (-210, 2167)	f: (772, 1478) u: (-105, 1132)
Page based Fixation Counts per Sequence based Fixation Counts	Familiar user have bigger proportion of sequence- based fixation counts over all fixations than unfamiliar user.	f: (16, 27) u: (7, 7)	f: (1.48, 0.31) u: (1.44, 0.28)	f: (2.65, 2.74) u: (4.03, 4.19)	f: (3.54, 2.61) u: (2.39, 0.73)	f: (1.27, 0.21) u: (1.48, 0.25)	f: (2.59, 1.61) u: (4.22, 2.37)

5.2 Descriptive Analysis

After the eye-tracking study is conducted again, descriptive analysis is done before the prediction. Based on the feature-level hypotheses (*see Tables 2.2 and 2.3*), the datasets are statistically analyzed and verified. This is a procedure which includes a series of statistical tests which show if the hypotheses are confirmed or not.

Because of the calibration problem of the mobile eye tracker, the participants are missing for certain web pages (*see Table* 5.2). Although the eye tracker recorded 13th and 16th participants' eye movements (*see Appendix*N), they are outliers for all gaze data; so, their data are not taken into account in descriptive analyses. Moreover, because Scanpath and First Fixated AOI values are String and Char, their t and w values cannot be calculated. They do not take place in descriptive analyses.

5.2.1 Descriptive Analysis of Familiarity

In order to investigate whether familiar and unfamiliar participants' eye movements are different or not, statistical tests are conducted [2]. They show significant differences and prove that the results are not by chance (*see Table* 5.4).

If the distributions of eye movement data for each group are normal, dependent T-Test (two-sided) is conducted. Otherwise, as a non-parametric test, the Wilcoxon Signed-Rank test is used. In order to show if the distribution is normal or not, the Shapiro-Wilk test is conducted. T-Test produces t value. In the t-test, the first value represents familiar, the second one belongs to unfamiliar participants. In other words, t values are interpreted in terms of familiar. For instance, if t value is high and positive, familiar's values are higher than unfamiliar; otherwise, unfamiliar's values are higher [39]. Wilcoxon Signed-Rank test produces w value. Hole shows the critical values of w value in the Wilcoxon Signed Rank test and we evaluate the results based on this [40].

The statistical tests were conducted with %95 confidence interval [39]. Thus, p-value should be lower than 0.05 threshold to prove that there is a statistically significant difference between familiar and unfamiliar. The t and w values with a star (*) mean that there is a significant difference. In other words, H0 is rejected and H1 (our hypothesis) is admitted. In addition, in order to show the strength of the differences between the two groups, The Cohen's d values were calculated [41]. The Cohen's d value is the effect size of which the first value belongs to familiar, the second one is for unfamiliar. The Cohen's d value was calculated as the effect size when both the dependent T-test and Wilcoxon were applied (.2: Small Effect, .5: Medium Effect, .8: Large Effect).

In addition to statistical tests, mean and standard deviations of eye-tracking datasets are examined (*see Table* 5.5). Those show the differences between familiar and unfamiliar's values. They are evaluated based on the familiarity hypothesis (H1).

Apple and Yahoo browsing dataset do not have any significant difference between familiar and unfamiliar. In browsing datasets, the BBC-browsing dataset has 2 differences which are significant and both are related to path angles. Searching datasets have more significant differences. According to Vicram framework, the most complex web page of this study has more significant differences. In the scope of this study, it is generalized that searching over a complex web page makes differences in the eye movements of familiar and unfamiliar users. Means and standard deviations support this inference (*see Table* 5.5).

Table 5.6: Statistical Analysis of Gender - Browsing

Features	Gender related Hypothesis (H1)	Browsing					
		Apple (t or w)	**AVG (t or w)**	**Babylon (t or w)**	**BBC (t or w)**	**GoDaddy (t or w)**	**Yahoo (t or w)**
Mean of Sequence based Fixation Durations	Female's fixation duration is longer than male	w: 6* d: -1.09 df: NA	w: 21 d: -0.49 df: NA	t: -0.65 d: -0.33 df: 9	t: 0.88 d: 0.45 df: 12	w: 30 d: -0.48 df: NA	t: -0.24 d: -0.12 df: 12
Sum of Sequence based Fixation Durations	Female's fixation duration is longer than male	w: 18 d: -0.39 df: NA	t: 0.19 d: 0.1 df: 12	t: 1.09 d: 0.51 df: 11	t: 0.32 d: 0.17 df: 9	t: 0.45 d: 0.23 df: 10	t: -0.38 d: -0.17 df: 13
Sequence based Fixation Counts	Male make fewer fixations than females	t: -0.08 d: -0.04 df: 10	t: 1.11 d: 0.55 df: 11	t: 1.55 d: 0.67 df: 14	t: -0.05 d: -0.02 df: 8	t: 0.71 d: 0.34 df: 13	t: -0.36 d: -0.16 df: 13
Percentage of First Fixated AOI	Duration of First Fixated AOI is different than other fixations as a percentage in terms of gender.	w: 19 d: -0.12 df: NA	w: 24 d: -0.43 df: NA	w: 25 d: -0.72 df: NA	t: -1.42 d: -0.69 df: 8	w: 32 d: -0.32 df: NA	w: 30 d: 0.49 df: NA
Duration of First Fixated AOI	Female's fixation duration on First Fixated AOI is longer than male	w: 14 d: -0.65 df: NA	w: 29 d: -0.41 df: NA	w: 27 d: -0.63 df: NA	t: -1.45 d: -0.70 df: 8	w: 28 d: 0.02 df: NA	t: 0.95 d: 0.61 df: 5
Mean of Page based Fixation Durations	Female's fixation duration is longer than male	t: -1.48* d: -0.84 df: 9	w: 24 d: -0.53 df: NA	t: -1.06 d: -0.5 df: 11	t: 0.62 d: 0.32 df: 12	t: -1.53 d: -0.67 df: 15	t: 0.72 d: 0.37 df: 10
Sum of Page based Fixation Durations	Female's fixation duration is longer than male	w: 18 d: -0.46 df: NA	t: 0.79 d: 0.39 df: 11	w: 50 d: 0.63 df: NA	t: -0.23 d: -0.12 df: 9	w: 37 d: -0.52 df: NA	t: 0.77 d: 0.38 df: 12
Page based Fixation Counts	Male make fewer fixations than females	t: 0.05 d: 0.02 df: 9	t: 1.2 d: 0.6 df: 11	t: 1.49 d: 0.73 df: 10	t: -0.43 d: -0.23 df: 9	t: 0.001 d: 0.0007 df: 9	t: -0.02 d: -0.01 df: 9
Number of Viewed AOIs per Page based Fixations	There are differences between familiar and unfamiliar user	t: -0.7 d: -0.4 df: 9	w: 29 d: -0.22 df: NA	t: -0.56 d: -0.33 df: 6	t: -1.36 d: -0.7 df: 12	t: 0.57 d: 0.29 df: 10	t: 1.75* d: 0.83 df: 13
Mean of Distances among Page based Fixations	Fixation distance differs in terms of gender.	t: 1.07 d: 0.57 df: 10	t: 1.9* d: 0.95 df: 12	t: -1.69 d: -0.65 df: 15	t: -0.86 d: -0.44 df: 13	t: 0.01 d: 0.007 df: 14	w: 41 d: 0.16 df: NA
Sum of Distances among Page based Fixations	Fixation distance differs in terms of gender.	t: 0.74 d: 0.39 df: 8	t: 1.99* d: 1.01 df: 12	t: -0.17 d: -0.07 df: 13	t: -0.92 d: -0.4 df: 12	t: -0.04 d: -0.02 df: 11	t: 0.31 d: 0.14 df: 13
Mean of Path Angles among Page based Fixations	Path angle differs in terms of gender.	w: 27 d: 0.54 df: NA	t: 0.17 d: 0.08 df: 11	t: 0.68 d: 0.35 df: 9	t: 0.84 d: 0.43 df: 12	t: -0.5 d: -0.24 df: 11	t: 0.66 d: 0.35 df: 9
Sum of Path Angles among Page based Fixations	Path angle differs in terms of gender.	t: 0.71 d: 0.4 df: 9	t: 0.74 d: 0.39 df: 8	t: 0.58 d: 0.31 df: 8	t: 1.0 d: 0.51 df: 12	t: -0.42 d: -0.21 df: 11	t: 0.54 d: 0.28 df: 10
Page based Fixation Counts per Sequence based Fixation Counts	Gender affects the proportion of looking AOI or non-AOI.	w: 23 d: -0.36 df: NA	t: -0.31 d: -0.15 df: 11	w: 29 d: -0.43 df: NA	t: -0.01 d: -0.008 df: 11	t: -1.45 d: -0.64 df: 15	w: 32 d: -0.48 df: NA

*p < 0.05

Table 5.7: Means and Standard Deviations of Gender - Browsing

Features	Gender related Hypothesis (H1)	Browsing					
		Apple (m, sd)	**AVG (m, sd)**	**Babylon (m, sd)**	**BBC (m, sd)**	**GoDaddy (m, sd)**	**Yahoo (m, sd)**
Mean of Sequence based Fixation Durations	Female's fixation duration is longer than male	f: (309, 79) m: (489, 227)	f: (350, 57) m: (538, 513)	f: (325, 41) m: (338, 39)	f: (367, 48) m: (345, 49)	f: (322, 25) m: (351, 72)	f: (340, 47) m: (347, 57)
Sum of Sequence based Fixation Durations	Female's fixation duration is longer than male	f: (4671, 2913) m: (5732, 2460)	f: (10736, 3184) m: (10338, 4552)	f: (19556, 4130) m: (17144, 4914)	f: (20212, 5385) m: (19450, 3332)	f: (20236, 6828) m: (18846, 5431)	f: (12489, 4852) m: (13819, 9038)
Sequence based Fixation Counts	Male make fewer fixations than females	f: (14, 8) m: (14, 8)	f: (30, 7) m: (25, 11)	f: (60, 9) m: (50, 15)	f: (56, 19) m: (56, 9)	f: (62, 19) m: (55, 19)	f: (37, 14) m: (41, 28)
Percentage of First Fixated AOI	Duration of First Fixated AOI is different than other fixations as a percentage in terms of gender.	f: (0.01, 0.008) m: (0.01, 0.009)	f: (0.01, 0.01) m: (0.04, 0.07)	f: (0.01, 0.002) m: (0.01, 0.01)	f: (0.009, 0.003) m: (0.01, 0.01)	f: (0.02, 0.02) m: (0.01, 0.008)	f: (0.01, 0.01) m: (0.01, 0.005)
Duration of First Fixated AOI	Female's fixation duration on First Fixated AOI is longer than male	f: (264, 84) m: (391, 272)	f: (469, 273) m: (884, 1340)	f: (300, 64) m: (498, 373)	f: (266, 77) m: (416, 279)	f: (478, 469) m: (471, 245)	f: (417, 366) m: (269, 127)
Mean of Page based Fixation Durations	Female's fixation duration is longer than male	f: (361, 62) m: (423, 83)	f: (349, 47) m: (458, 275)	f: (330, 38) m: (352, 46)	f: (364, 44) m: (349, 44)	f: (332, 44) m: (373, 69)	f: (388, 50) m: (369, 49)
Sum of Page based Fixation Durations	Female's fixation duration is longer than male	f: (24871, 8259) m: (27789, 1836)	f: (26324, 3698) m: (24189, 6511)	f: (28636, 844) m: (27115, 2817)	f: (27079, 2373) m: (27322, 1402)	f: (25636, 7136) m: (28022, 1405)	f: (27301, 2635) m: (26169, 3128)
Page based Fixation Counts	Male make fewer fixations than females	f: (68, 22) m: (67, 12)	f: (76, 16) m: (63, 26)	f: (87, 12) m: (78, 13)	f: (76, 16) m: (79, 10)	f: (77, 21) m: (77, 14)	f: (71, 12) m: (71, 10)
Number of Viewed AOIs per Page based Fixations	There are differences between familiar and unfamiliar user	f: (0.06, 0.02) m: (0.07, 0.03)	f: (0.06, 0.03) m: (0.07, 0.04)	f: (0.1, 0.03) m: (0.1, 0.01)	f: (0.07, 0.02) m: (0.09, 0.02)	f: (0.08, 0.02) m: (0.07, 0.01)	f: (0.09, 0.02) m: (0.06, 0.03)
Mean of Distances among Page based Fixations	Fixation distance differs in terms of gender.	f: (196, 49) m: (171, 32)	f: (184, 26) m: (152, 38)	f: (162, 17) m: (187, 44)	f: (176, 49) m: (200, 57)	f: (180, 31) m: (180, 37)	f: (194, 34) m: (184, 73)
Sum of Distances among Page based Fixations	Fixation distance differs in terms of gender.	f: (13207, 5781) m: (11391, 2710)	f: (14064, 3764) m: (9760, 4602)	f: (14258, 3325) m: (14576,4414)	f: (13320, 4944) m: (15656, 4786)	f: (13788, 4962) m: (13891, 4436)	f: (13708, 3164) m: (13034, 5364)
Mean of Path Angles among Page based Fixations	Path angle differs in terms of gender.	f: (11, 11) m: (5, 12)	f: (6, 19) m: (4, 15)	f: (-1.7, 11) m: (-5, 10)	f: (-3, 7) m: (-6, 9)	f: (-0.7, 10) m: (1.7, 9)	f: (-7, 11) m: (-11, 10)
Sum of Path Angles among Page based Fixations	Path angle differs in terms of gender.	f: (617, 606) m: (326, 826)	f: (654, 1567) m: (161, 840)	f: (-169, 979) m: (-437, 794)	f: (-178, 729) m: (-558, 737)	f: (-23, 811) m: (130, 670)	f: (-631, 838) m: (-866, 823)
Page based Fixation Counts per Sequence based Fixation Counts	Gender affects the proportion of looking AOI or non-AOI.	f: (8, 7) m: (15, 27)	f: (3, 0.7) m: (3, 1)	f: (1.4, 0.2) m: (1.6, 0.3)	f: (1.4, 0.3) m: (1.4, 0.2)	f: (1.2, 0.2) m: (1.5, 0.3)	f: (2.2, 0.9) m: (3.8, 4.1)

Table 5.8: Statistical Analysis of Gender - Searching

Features	Gender related Hypothesis (H1)	Searching					
		Apple (t or w)	AVG (t or w)	Babylon (t or w)	BBC (t or w)	GoDaddy (t or w)	Yahoo (t or w)
Mean of Sequence based Fixation Durations	Female's fixation duration is longer than male	w: 36 d: 0.25 df: NA	w: 43 d: 0.73 df: NA	t: 1.11 d: 0.63 df: 7	w: 14 d: -0.21 df: NA	w: 28 d: -0.13 df: NA	t: -0.002 d: -0.001 df: 11
Sum of Sequence based Fixation Durations	Female's fixation duration is longer than male	t: 1.11 d: 0.67 df: 6	t: 1.18 d: 0.68 df: 7	t: 0.6 d: 0.29 df: 11	t: -0.33 d: -0.2 df: 5	t: 0.97 d: 0.43 df: 13	t: 2.29* d: 1.49 df: 5
Sequence based Fixation Counts	Male make fewer fixations than females	t: 1.55* d: 0.98 df: 5	t: -0.34 d: -0.19 df: 7	t: 0.19 d: 0.09 df: 13	w: 10 d: -0.13 df: NA	t: 1.25 d: 0.58 df: 13	w: 49* d: 1.43 df: NA
Percentage of First Fixated AOI	Duration of First Fixated AOI is different than other fixations as a percentage in terms of gender.	w: 14 d: -0.63 df: NA	w: 35 d: 0.69 df: NA	t: -1.46 d: -0.62 df: 15	w: 11 d: 0.08 df: NA	w: 27 d: -0.5 df: NA	t: -1.77* d: -0.86 df: 9
Duration of First Fixated AOI	Female's fixation duration on First Fixated AOI is longer than male	w: 28 d: -0.35 df: NA	w: 40 d: 0.71 df: NA	w: 27 d: -0.43 df: NA	t: -0.71 d: -0.43 df: 5	w: 29 d: -0.42 df: NA	t: 0.07 d: 0.05 df: 4
Mean of Page based Fixation Durations	Female's fixation duration is longer than male	w: 30 d: 0.007 df: NA	w: 48 d: 0.74 df: NA	t: 1.03 d: 0.6 df: 6	w: 14 d: -0.04 df: NA	w: 30 d: -0.01 df: NA	w: 33 d: -0.16 df: NA
Sum of Page based Fixation Durations	Female's fixation duration is longer than male	t: 1.58* d: 0.97 df: 6	t: 0.98 d: 0.46 df: 12	t: 0.17 d: 0.08 df: 13	t: -0.17 d: -0.11 df: 7	t: 0.87 d: 0.48 df: 8	t: 2.55* d: 1.48 df: 6
Page based Fixation Counts	Male make fewer fixations than females	t: 1.76* d: 1.18 df: 5	t: -0.12 d: -0.07 df: 6	t: -0.27 d: -0.13 df: 13	t: 0.02 d: 0.02 df: 6	t: 0.84 d: 0.48 df: 7	t: 2.61* d: 1.38 df: 7
Number of Viewed AOIs per Page based Fixations	There are differences between familiar and unfamiliar user	w: 21 d: -0.1 df: NA	w: 35 d: 0.79 df: NA	t: 0.51 d: 0.23 df: 15	t: 0.22 d: 0.15 df: 5	t: -0.97 d: -0.47 df: 12	w: 14 d: -0.72 df: NA
Mean of Distances among Page based Fixations	Fixation distance differs in terms of gender.	w: 17 d: -0.58 df: NA	t: 0.94 d: 0.58 df: 6	t: -0.95 d: -0.44 df: 14	w: 2* d: -1.26 df: NA	t: 0.47 d: 0.26 df: 8	t: 1.72* d: 0.97 df: 6
Sum of Distances among Page based Fixations	Fixation distance differs in terms of gender.	t: 0.98 d: 0.67 df: 5	t: -0.23 d: -0.12 df: 8	t: -0.68 d: -0.31 df: 15	w: 8 d: -0.09 df: NA	t: 1.03 d: 0.57 df: 8	w: 54* d: 2.13 df: NA
Mean of Path Angles among Page based Fixations	Path angle differs in terms of gender.	t: 0.66 d: 0.37 df: 7	w: 39 d: -0.47 df: NA	t: -0.83 d: -0.45 df: 8	t: 2.10* d: 1.32 df: 6	t: -0.13 d: -0.06 df: 11	t: 0.87 d: 0.48 df: 6
Sum of Path Angles among Page based Fixations	Path angle differs in terms of gender.	t: 1.01 d: 0.73 df: 4	t: 1.36 d: 0.56 df: 15	t: -0.57 d: -0.3 df: 9	w: 18* d: 1.09 df: NA	t: 0.57 d: 0.33 df: 7	t: 1.01 d: 0.76 df: 4
Page based Fixation Counts per Sequence based Fixation Counts	Gender affects the proportion of looking AOI or non-AOI.	w: 23 d: -0.3 df: NA	t: -0.2 d: -0.1 df: 8	w: 21 d: -0.52 df: NA	t: -0.41 d: -0.28 df: 5	w: 28 d: -0.45 df: NA	t: -1.13 d: -0.46 df: 14

*p < 0.05

Table 5.9: Means and Standard Deviations of Gender - Searching

Features	Gender related Hypothesis (H1)	Searching					
		Apple (m, sd)	AVG (m, sd)	Babylon (m, sd)	BBC (m, sd)	GoDaddy (m, sd)	Yahoo (m, sd)
Mean of Sequence based Fixation Durations	Female's fixation duration is longer than male	f: (375, 77) m: (348, 119)	f: (394, 54) m: (368, 47)	f: (404, 92) m: (363, 39)	f: (392, 83) m: (444, 302)	f: (376, 91) m: (378, 79)	f: (352, 41) m: (352, 60)
Sum of Sequence based Fixation Durations	Female's fixation duration is longer than male	f: (18246, 11797) m: (11579, 8874)	f: (20212, 8611) m: (15612, 5709)	f: (47601, 21326) m: (41656, 18957)	f: (23413, 9663) m: (26945, 20701)	f: (101948, 24783) m: (85034, 44606)	f: (25900, 11300) m: (13390, 7017)
Sequence based Fixation Counts	Male make fewer fixations than females	f: (47, 22) m: (30, 14)	f: (39, 21) m: (42, 15)	f: (118, 46) m: (113, 49)	f: (65, 39) m: (69, 62)	f: (277, 71) m: (220, 110)	f: (74, 33) m: (39, 20)
Percentage of First Fixated AOI	Duration of First Fixated AOI is different than other fixations as a percentage in terms of gender.	f: (0.007, 0.003) m: (0.01, 0.01)	f: (0.07, 0.1) m: (0.01, 0.007)	f: (0.006, 0.002) m: (0.008, 0.004)	f: (0.01, 0.01) m: (0.01, 0.005)	f: (0.003, 0.002) m: (0.01, 0.02)	f: (0.006, 0.005) m: (0.01, 0.006)
Duration of First Fixated AOI	Female's fixation duration on First Fixated AOI is longer than male	f: (269, 54) m: (356, 291)	f: (209, 42) m: (367, 253)	f: (374, 211) m: (481, 261)	f: (304, 166) m: (463, 185)	f: (364, 98) m: (408, 124)	f: (473, 380) m: (460, 182)
Mean of Page based Fixation Durations	Female's fixation duration is longer than male	f: (376, 54) m: (375, 107)	f: (388, 73) m: (365, 61)	f: (416, 117) m: (368, 40)	f: (385, 90) m: (391, 166)	f: (389, 83) m: (390, 59)	f: (427, 81) m: (459, 217)
Sum of Page based Fixation Durations	Female's fixation duration is longer than male	f: (41395, 17192) m: (27698, 12500)	f: (42897, 10420) m: (37368, 12612)	f: (64437, 23920) m: (62380, 25660)	f: (31878, 19680) m: (34454, 25530)	f: (108103, 49750) m: (87790, 36244)	f: (76432, 23808) m: (45738, 19345)
Page based Fixation Counts	Male make fewer fixations than females	f: (111, 45) m: (72, 24)	f: (100, 56) m: (103, 32)	f: (161, 63) m: (169, 66)	f: (94, 78) m: (93, 70)	f: (286, 146) m: (229, 98)	f: (179, 50) m: (108, 51)
Number of Viewed AOIs per Page based Fixations	There are differences between familiar and unfamiliar user	f: (0.06, 0.03) m: (0.06, 0.02)	f: (0.09, 0.12) m: (0.03, 0.01)	f: (0.08, 0.02) m: (0.07, 0.02)	f: (0.11, 0.06) m: (0.1, 0.04)	f: (0.03, 0.01) m: (0.04, 0.01)	f: (0.04, 0.01) m: (0.06, 0.04)
Mean of Distances among Page based Fixations	Fixation distance differs in terms of gender.	f: (154, 42) m: (175, 33)	f: (196, 53) m: (174, 28)	f: (146, 27) m: (160, 32)	f: (151, 33) m: (186, 21)	f: (180, 43) m: (170, 34)	f: (158, 37) m: (126, 32)
Sum of Distances among Page based Fixations	Fixation distance differs in terms of gender.	f: (18486,11848) m: (12966, 5912)	f: (17297, 9068) m: (18321, 7870)	f: (23982, 9804) m: (27577, 12462)	f: (15516, 14634) m: (16827, 13140)	f: (50891, 24672) m: (38877, 18598)	f: (28210, 8500) m: (13320, 6318)
Mean of Path Angles among Page based Fixations	Path angle differs in terms of gender.	f: (6.25, 10.54) m: (2.48, 9.73)	f: (-12.9, 31.9) m: (-3.17, 12.53)	f: (-2.05, 14.73) m: (3.04, 8.35)	f: (5.79, 6.13) m: (-5.87, 10.34)	f: (5.81, 7.52) m: (6.33, 8.09)	f: (7.17, 12.4) m: (1.59, 11.12)
Sum of Path Angles among Page based Fixations	Path angle differs in terms of gender.	f: (825, 1493) m: (127, 562)	f: (82, 630) m: (-490, 1147)	f: (-209, 2275) m: (354, 1502)	f: (646, 1038) m: (-716, 1379)	f: (1750, 2284) m: (1161, 1390)	f: (1255, 2254) m: (197, 857)
Page based Fixation Counts per Sequence based Fixation Counts	Gender affects the proportion of looking AOI or non-AOI.	f: (2.53, 0.88) m: (3.12, 2.24)	f: (2.47, 0.88) m: (2.55, 0.73)	f: (1.37, 0.37) m: (1.61, 0.48)	f: (1.32, 0.3) m: (1.39, 0.22)	f: (0.98, 0.24) m: (1.21, 0.62)	f: (2.73, 1.16) m: (3.71, 2.36)

5.2.2 Descriptive Analysis of Gender

The same statistical tests with the familiarity are conducted for genders (*see Section* 5.2.1). In browsing datasets, t and w values are calculated and presented (*see Table* 5.6). The Cohen's d value shows the effect size of the significance which is two-tailed. The first value belongs to Female while the second one represents Male. In browsing datasets, Apple and AVG have 2 significant differences while Yahoo has one significant difference. The significant differences in the Apple-browsing dataset are on the Mean of Sequence-based and Page-based Fixation Durations. According to Vicram Framework, Apple's complexity is low. AVG has significant differences in the Mean and Sum of Distances among Page-based Fixations. AVG's complexity is medium. Lastly, Yahoo has a significant difference in the Number of Viewed AOIs per Page-based Fixation Counts. Yahoo's complexity is also medium. Mean and standard deviation values support these differences (*see Table* 5.7).

In searching datasets, AVG (Medium), Babylon (Low) and GoDaddy (High) have no significant difference which does not depend on visual complexity in this study because each one has different visual complexity levels based on Vicram Framework. However, searching datasets have more significant differences. Apple has two values which are on Sequence-based Fixation Counts and Sum of Page based Fixation Durations. BBC has three significant differences in the Mean of Distances among Page based Fixations, Mean and Sum of Path Angles among Page based Fixations. Yahoo has seven significant differences which are on the features; Sum of Sequence-based and Page based Fixation Durations, Sequence-based and Page based Fixation Counts, Mean and Sum of Distances among Page based Fixations, and Percentage of First Fixated AOI. Therefore, in the scope of this study, eye movement data implies gender better on searching data and page-based values.

5.3 Validation Results

The validation study aims at verifying data models and testing data models with independent datasets. Data models are validated in two ways. Firstly, 20 users eye tracking data is added to the existing 79 users' data and data modeling procedure is repeated again. Secondly, the data models try to predict the familiarity and gender of the new datasets independently. Validation results are presented similar to modeling results. Accuracy, precision, recall and F-measure of data models and tests are presented.

Table 5.10: Validated Familiarity Models by Logistic Regression

		Browsing			Searching		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	67.02%	57.29%	44.32%	69.47%	44.94%	55.10%
	Precision	0.614	0.58	0.296	0.669	0.383	0.496
	Recall	0.670	0.573	0.443	0.695	0.449	0.551
	F-measure	0.605	0.557	0.355	0.655	0.38	0.463
	Number of Instances	94	96	97	95	89	98
Resampling	**Accuracy**	90.42%	73.95%	83.50%	72.63%	86.51%	84.69%
	Precision	0.909	0.752	0.872	0.743	0.879	0.856
	Recall	0.904	0.740	0.835	0.726	0.865	0.847
	F-measure	0.900	0.735	0.825	0.732	0.865	0.843
	Number of Instances	94	96	97	95	89	98
SMOTE	**Accuracy**	88.07%	84.87%	86.32%	90.00%	86.36%	88.23%
	Precision	0.883	0.848	0.863	0.902	0.864	0.883
	Recall	0.881	0.849	0.863	0.900	0.864	0.882
	F-measure	0.881	0.848	0.863	0.900	0.865	0.882
	Number of Instances	109	119	117	110	110	119

Table 5.11: Validated Familiarity Models by SMO

		Browsing			Searching		
	Algorithms	**Apple**	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	68.08%	55.20%	55.67%	67.36%	47.19	58.16%
	Precision	0.681	0.565	0.328	0.674	0.393	0.617
	Recall	0.681	0.552	0.557	0.674	0.472	0.582
	F-measure	0.810	0.516	0.413	0.805	0.383	0.454
	Number of Instances	94	96	97	95	89	98
Resampling	**Accuracy**	90.42%	70.83%	82.47%	86.31%	83.14%	85.71%
	Precision	0.916	0.709	0.866	0.886	0.831	0.886
	Recall	0.904	0.708	0.825	0.863	0.831	0.857
	F-measure	0.899	0.708	0.813	0.851	0.831	0.851
	Number of Instances	94	96	97	95	89	98
SMOTE	**Accuracy**	72.47%	76.47%	75.21%	71.82%	76.36%	76.47%
	Precision	0.813	0.765	0.823	0.810	0.774	0.831
	Recall	0.725	0.765	0.752	0.718	0.764	0.765
	F-measure	0.682	0.765	0.741	0.674	0.765	0.756
	Number of Instances	109	119	117	110	110	119

5.3.1 Familiarity Data Models Validation

Tables 5.10 and 5.11 show the results of repeated familiarity data models with the new datasets. In raw data models, the results are 2% worse than modeling results as average. On the other hand, resampled and oversampled validation data models produce better results than modeling as expected because the more the number of instances increases, the better the data models are trained.

In the raw data models, searching and browsing datasets produce similar results. However, the SMO algorithm trains better than Logistic Regression. In terms of visual complexity, the lowest complex web page, Apple has the best results. The lowest accuracy in the raw data models is 44.32% which belongs to Yahoo browsing data trained by Logistic Regression. The highest accuracy value is 69.47% which belongs to Apple searching data trained by Logistic Regression. Precision and recall measures are balanced for all raw data models.

In the resampled data models, searching datasets produce 3% better results than browsing datasets as average. Moreover, the SMO algorithm trains 3% better than Logistic Regression as average. In terms of visual complexity, there is no clear evidence to say the visual complexity influences familiarity prediction in resampled datasets. The lowest accuracy in the resampled data models is 70.83% which belongs to BBC browsing data trained by SMO. The highest accuracy value is 90.42% which belongs to Apple browsing data trained by both Logistic Regression and SMO algorithms. Precision and recall measures are balanced for all resampled data models.

In the oversampled data models, searching datasets produce 1% better results than browsing datasets as average. However, Logistic Regression trains 13% better than the SMO algorithm as average. In terms of visual complexity, there is no clear evidence to say the visual complexity influences familiarity prediction in resampled datasets. The lowest accuracy in the synthetic oversampled data models is 71.82% which belongs to Apple searching data trained by the SMO algorithm. The highest accuracy value is 90.00% which belongs to Apple searching data trained by Logistic Regression. Precision and recall measures are balanced for all oversampled data models.

Table 5.12: Validated Gender Models by Logistic Regression with Browsing Data

		Browsing					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	44.56%	53.19%	52.04%	62.50%	47.95%	45.36%
	Precision	0.447	0.535	0.527	0.639	0.481	0.448
	Recall	0.447	0.532	0.520	0.625	0.480	0.454
	F-measure	0.447	0.521	0.507	0.618	0.469	0.449
	Number of Instances	94	94	98	96	98	97
Resampling	**Accuracy**	71.27%	74.46%	73.46%	88.54%	73.46%	77.31%
	Precision	0.713	0.746	0.739	0.887	0.742	0.788
	Recall	0.713	0.745	0.735	0.885	0.735	0.773
	F-measure	0.713	0.744	0.734	0.885	0.733	0.772
	Number of Instances	94	94	98	96	98	97
SMOTE	**Accuracy**	73.50%	80.34%	77.04%	78.15%	75.40%	73.33%
	Precision	0.735	0.802	0.770	0.786	0.752	0.732
	Recall	0.735	0.803	0.770	0.782	0.754	0.733
	F-measure	0.725	0.801	0.767	0.775	0.752	0.732
	Number of Instances	117	117	122	119	122	120

Table 5.13: Validated Gender Models by SMO with Browsing Data

		Browsing					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	43.61%	50.00%	47.95%	51.04%	57.14%	39.17%
	Precision	0.436	0.500	0.478	0.509	0.573	0.369
	Recall	0.436	0.500	0.480	0.511	0.571	0.392
	F-measure	0.436	0.500	0.478	0.503	0.566	0.371
	Number of Instances	94	94	98	96	98	97
Resampling	**Accuracy**	76.59%	80.85%	74.48%	89.58%	74.48%	84.53%
	Precision	0.774	0.83	0.745	0.898	0.746	0.852
	Recall	0.766	0.809	0.745	0.896	0.745	0.845
	F-measure	0.764	0.805	0.745	0.794	0.744	0.844
	Number of Instances	94	94	98	96	98	97
SMOTE	**Accuracy**	67.52%	70.08%	66.39%	73.10%	68.03%	65.00%
	Precision	0.669	0.708	0.659	0.728	0.675	0.654
	Recall	0.675	0.701	0.664	0.731	0.680	0.650
	F-measure	0.662	0.677	0.659	0.728	0.674	0.651
	Number of Instances	117	117	122	119	122	120

Table 5.14: Validated Gender Models by Logistic Regression with Searching Data

		Searching					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	37.89%	50.00%	61.22%	52.80%	54.16%	54.08%
	Precision	0.363	0.502	0.612	0.534	0.548	0.539
	Recall	0.379	0.500	0.612	0.528	0.542	0.541
	F-measure	0.364	0.495	0.612	0.496	0.525	0.533
	Number of Instances	95	98	98	89	96	98
Resampling	**Accuracy**	78.94%	74.48%	74.48%	75.28%	79.16%	81.63%
	Precision	0.797	0.759	0.772	0.758	0.792	0.820
	Recall	0.789	0.745	0.745	0.753	0.792	0.816
	F-measure	0.787	0.742	0.740	0.752	0.792	0.815
	Number of Instances	95	98	98	89	96	98
SMOTE	**Accuracy**	72.03%	73.77%	81.96%	76.57%	80.00%	79.33%
	Precision	0.723	0.748	0.821	0.774	0.804	0.820
	Recall	0.720	0.738	0.820	0.766	0.800	0.793
	F-measure	0.721	0.740	0.820	0.768	0.801	0.794
	Number of Instances	118	122	122	111	120	121

Table 5.15: Validated Gender Models by SMO with Searching Data

		Searching					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	36.84%	53.06%	67.34%	51.68%	53.12%	56.12%
	Precision	0.329	0.532	0.692	0.517	0.537	0.593
	Recall	0.368	0.531	0.673	0.517	0.531	0.561
	F-measure	0.334	0.508	0.663	0.502	0.512	0.496
	Number of Instances	95	98	98	89	96	98
Resampling	**Accuracy**	78.94%	73.46%	77.55%	77.52%	77.08%	84.69%
	Precision	0.791	0.735	0.782	0.790	0.779	0.882
	Recall	0.789	0.735	0.776	0.775	0.771	0.847
	F-measure	0.789	0.735	0.774	0.773	0.769	0.842
	Number of Instances	95	98	98	89	96	98
SMOTE	**Accuracy**	65.25%	70.49%	76.22%	67.56%	69.16%	71.90%
	Precision	0.673	0.712	0.774	0.692	0.689	0.752
	Recall	0.653	0.705	0.762	0.676	0.692	0.719
	F-measure	0.655	0.707	0.764	0.679	0.675	0.719
	Number of Instances	118	122	122	111	120	121

5.3.2 Gender Data Models Validation

Tables 5.12, 5.13, 5.14, and 5.15 show the results of repeated gender data models with the new datasets. The accuracy values of new data models are 3% better than previous data models as average. Searching data models are more predictive than browsing ones.

In the raw data models, searching values are better than browsing and in browsing data, Logistic Regression produces 2% better results than SMO while in searching data, SMO produces 2% better results than LR as average. In terms of visual complexity, there is no consistent evidence to claim that visual complexity is an influencer while predicting gender. The lowest accuracy in the raw data models is 36.84% which belongs to Apple searching data trained by SMO. The highest accuracy value is 67.34% which belongs to Babylon searching data trained by SMO. Precision and recall measures are balanced for all raw data models.

In the resampled data models, searching values are almost equal to browsing data models. SMO produces 3% better results than Logistic Regression as average. The results do not show a pattern to claim that visual complexity is an influencer while predicting gender in the resampled dataset. The lowest accuracy value in the resampled data models is 71.27% which belongs to Apple browsing data trained by Logistic Regression. The highest accuracy value is 89.58% belongs to BBC browsing data trained by SMO. Precision and recall measures are balanced for all resampled data models.

In the synthetic oversampled data models, searching results are 1% better than browsing results as average. SMO produces 7% better results than Logistic Regression as average. Visual complexity level does not influence training eye movement data. The lowest accuracy value in the oversampled data models is 65.00% which belongs to Yahoo browsing data trained by SMO. The highest accuracy value is 81.96% belongs to Babylon searching data trained by Logistic Regression. Precision and recall measures are balanced for all oversampled data models.

5.3.3 Familiarity Prediction on Test Set

Tables 5.16 and 5.17 show the results of predicting familiarity factor of the new datasets. The accuracy values are lower than the expected for resampled and oversampled data models. Moreover, recall values of all resampled and oversampled data models are 1.00 which show the prediction is biased. Precision and recall need to be balanced. In raw data models, Logistic Regression produces the best results in terms of both accuracy values and precision-recall balance.

In the raw data models, browsing data models by Logistic Regression predict the

Table 5.16: Familiarity Prediction by Logistic Regression

		Browsing			Searching		
	Algorithms	**Apple**	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	53.33%	47.05%	50.00%	43.75%	30.00%	36.84%
	Precision	0.589	0.265	0.344	0.233	0.133	0.368
	Recall	0.533	0.471	0.500	0.438	0.300	1.00
	F-measure	0.489	0.339	0.407	0.304	0.185	0.538
	Number of Instances	15	17	18	16	10	19
Resampling	**Accuracy**	53.33%	52.94%	61.11%	50.00%	40.00%	63.15%
	Precision	0.533	0.529	0.611	0.500	0.400	0.632
	Recall	1.00	1.00	1.00	1.000	1.00	1.00
	F-measure	0.696	0.692	0.759	0.667	0.571	0.774
	Number of Instances	15	17	18	16	10	19
SMOTE	**Accuracy**	46.66%	52.94%	61.11%	50.00%	40.00%	63.15%
	Precision	0.467	0.529	0.611	0.500	0.400	0.632
	Recall	1.00	1.00	1.00	1.000	1.00	1.00
	F-measure	0.636	0.692	0.759	0.667	0.571	0.774
	Number of Instances	15	17	18	16	10	19

Table 5.17: Familiarity Prediction by SMO

		Browsing			Searching		
	Algorithms	**Apple**	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual Complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	53.33%	47.05%	33.33%	50.00%	60.00%	36.84%
	Precision	0.533	0.471	0.137	0.500	0.600	0.368
	Recall	1.00	1.00	0.333	1.00	1.00	1.00
	F-measure	0.696	0.640	0.194	0.667	0.750	0.538
	Number of Instances	15	17	18	16	10	19
Resampling	**Accuracy**	53.33%	52.94%	61.11%	50.00%	40.00%	63.15%
	Precision	0.533	0.529	0.611	0.500	0.400	0.632
	Recall	1.00	1.00	1.00	1.000	1.00	1.00
	F-measure	0.696	0.692	0.759	0.667	0.571	0.774
	Number of Instances	15	17	18	16	10	19
SMOTE	**Accuracy**	46.66%	52.94%	61.11%	50.00%	40.00%	63.15%
	Precision	0.467	0.529	0.611	0.500	0.400	0.632
	Recall	1.00	1.00	1.00	1.000	1.00	1.00
	F-measure	0.636	0.692	0.759	0.667	0.571	0.774
	Number of Instances	15	17	18	16	10	19

best. Its worst accuracy value is 47.05% which belongs to BBC data model. The best accuracy value is 53.33% which belongs to Apple data model. The accuracy values of resampled and oversampled data are higher than raw data models and 50% threshold; but, their precision and recall values are imbalanced.

Table 5.18: Gender Prediction by Logistic Regression - Browsing

		Browsing					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	33.33%	43.75%	40.00%	52.94%	50.00%	38.88%
	Precision	0.179	0.391	0.433	0.631	0.500	0.472
	Recall	0.333	0.438	0.400	0.529	0.500	0.389
	F-measure	0.233	0.401	0.400	0.499	0.500	0.398
	Number of Instances	15	16	20	17	20	18
Resampling	**Accuracy**	46.66%	43.75%	40.00%	41.17%	40.00%	33.33%
	Precision	0.467	0.438	0.400	0.412	0.400	0.333
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.636	0.609	0.571	0.583	0.571	0.500
	Number of Instances	15	16	20	17	20	18
SMOTE	**Accuracy**	53.33%	56.25%	60.00%	58.82%	60.00%	66.66%
	Precision	0.533	0.563	0.600	0.588	0.600	0.667
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.696	0.720	0.750	0.741	0.750	0.800
	Number of Instances	15	16	20	17	20	18

Table 5.19: Gender Prediction by SMO - Browsing

		Browsing					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	60.00%	43.75%	40.00%	47.05%	50.00%	27.77%
	Precision	0.600	0.438	0.433	0.569	0.500	0.344
	Recall	0.600	1.00	0.400	0.471	0.500	0.278
	F-measure	0.589	0.609	0.400	0.416	0.500	0.262
	Number of Instances	15	16	20	17	20	18
Resampling	**Accuracy**	46.66%	43.75%	40.00%	41.17%	40.00%	33.33%
	Precision	0.467	0.438	0.400	0.412	0.400	0.333
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.636	0.609	0.571	0.583	0.571	0.500
	Number of Instances	15	16	20	17	20	18
SMOTE	**Accuracy**	53.33%	56.25%	60.00%	58.82%	60.00%	66.66%
	Precision	0.533	0.563	0.600	0.588	0.600	0.667
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.696	0.720	0.750	0.741	0.750	0.800
	Number of Instances	15	16	20	17	20	18

Table 5.20: Gender Prediction by Logistic Regression - Searching

		Searching					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	56.25%	45.00%	40.00%	50.00%	55.55%	63.15%
	Precision	0.375	0.565	0.400	0.476	0.556	0.632
	Recall	0.563	0.450	1.00	0.500	1.00	1.00
	F-measure	0.450	0.384	0.571	0.484	0.714	0.774
	Number of Instances	16	20	20	10	18	19
Resampling	**Accuracy**	37.50%	40.00%	40.00%	40.00%	55.55%	36.84%
	Precision	0.375	0.400	0.400	0.400	0.556	0.368
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.545	0.571	0.571	0.571	0.714	0.538
	Number of Instances	16	20	20	10	18	19
SMOTE	**Accuracy**	62.50%	60.00%	60.00%	60.00%	55.55%	63.15%
	Precision	0.625	0.600	0.600	0.600	0.556	0.632
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.769	0.750	0.750	0.750	0.714	0.774
	Number of Instances	16	20	20	10	18	19

Table 5.21: Gender Prediction by SMO - Searching

		Searching					
		Apple	AVG	Babylon	BBC	GoDaddy	Yahoo
	Visual Compexity	Low	Medium	Low	High	High	Medium
Raw Data	**Accuracy**	43.75%	40.00%	40.00%	50.00%	55.55%	63.15%
	Precision	0.337	0.456	0.400	0.476	0.556	0.632
	Recall	0.438	0.400	1.00	0.500	1.00	1.00
	F-measure	0.380	0.301	0.571	0.484	0.714	0.774
	Number of Instances	16	20	20	10	18	19
Resampling	**Accuracy**	37.50%	40.00%	40.00%	40.00%	55.55%	36.84%
	Precision	0.375	0.400	0.400	0.400	0.556	0.368
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.545	0.571	0.571	0.571	0.714	0.538
	Number of Instances	16	20	20	10	18	19
SMOTE	**Accuracy**	62.50%	60.00%	60.00%	60.00%	55.55%	63.15%
	Precision	0.625	0.600	0.600	0.600	0.556	0.632
	Recall	1.00	1.00	1.00	1.00	1.00	1.00
	F-measure	0.769	0.750	0.750	0.750	0.714	0.774
	Number of Instances	16	20	20	10	18	19

5.3.4 Gender Prediction on Test Set

Tables 5.18, 5.19, 5.20, and 5.21 show the results of predicting gender factor of the new datasets. Raw data models are balanced in terms of precision and recall while resampled and oversampled data models are imbalanced. Searching data models predict 6% better than browsing data models as average. Moreover, Logistic Regression predicts better than SMO. In raw data models, the lowest accuracy value is 27.77% which belongs to Yahoo browsing data model trained by SMO. The highest accuracy value is 63.15% which belongs to Yahoo searching data models trained by both algorithms.

5.4 Baseline Analysis for Predictions

Predictions of data models on test dataset for both gender and familiarity are unexpectedly bad. Thus, it needs to show correct predictions for familiarity and gender.

In familiarity, Appendix K shows actual values and the predictions by both logistic regression and SMO on raw, resampled and oversampled data. It is understood from the tables that while logistic regression predicts more familiar on raw data, SMO predicts unfamiliar correctly. On the resampled data, logistic regression and SMO predicts more unfamiliar although, on the oversampled data, both classifiers predict more familiar correctly. Except for the Yahoo-searching case, both classifiers predict resampled and oversampled data as either familiar or unfamiliar in %100. In Yahoo-searching, the logistic regression classifier predicts a resampled instance as familiar when the rest is predicted as unfamiliar in %95. However, it is unfamiliar in actual.

In gender, Appendix L presents tables which include actual values and the predictions of the classifiers. On the raw data, logistic regression classifies males more correctly than females and SMO also classifies males more correctly. On resampled and oversampled, both classifiers predict males better than females similarly. Except for BBC-browsing, logistic regression and SMO predict resampled and oversampled data as either female or male in %100. In the BBC-browsing case, logistic regression predicts females in %82 and males in %18. According to actual values, 2 of 3 are predicted as males correctly.

5.5 Summary and Conclusion

In this chapter, Section 5.1 explains essences of the validation study in detail. In the scope of this study, Eraslan's eye tracking study is followed and repeated in Ankara with 20 users again [1, 2]. Validation is conducted in two ways; adding new dataset to existing one to create validated data models and predicting new dataset's gender and familiarity by using existing data models.

Firstly, validated data models produce better accuracy values than the existing ones. It is important to see better accuracy values when number of instances increase as expected. Secondly, while predicting factors of new datasets, the existing resampled and oversampled data models are unsuccessful because their accuracy values are lower than the existing data models' accuracies and their recall values are 1.00 unexpectedly which means that prediction is random. Raw data models' precision-recall values are balanced; but their accuracy values are very low to make predictions.

When the training and testing datasets' descriptive analysis are conducted, it is seen that both data distributions differ from each other. This may be originated from research limitations which will be discussed in Chapter 6.

CHAPTER 6

CONCLUSION

The Internet has become a part of our daily lives. People spend a considerable part of the day for entertaining, training, or even working via the Internet. The web is a critical and very important application on the Internet. Avoiding the negative effects of the traditional "one-size-fits-all" approach is to develop web pages which adapt the behaviors and features of individual users and groups of users [42]. The overall objectives of this thesis enable adaptation of web pages to the users' gender and familiarity to the web page. In order to detect users' gender and familiarity, their eye movement data on 6 web pages is recorded and analyzed by data mining methods.

This thesis hypothesizes that eye-tracking data can be classified by data mining methods to predict a user's familiarity and gender. In order to test this hypothesis, a two-stage study is conducted; modeling and validation. Before modeling and validation, possible eye gaze features such as fixation duration, saccade length and so on are identified in the literature. The dataset consists of 16 independent eye tracking variables and a dependent user characteristic variable; familiarity or gender. There are mainly two types of eye gaze features in this study; sequence-based and page-based features. Sequence-based features are calculated with fixations over Areas of Interest (AOI) while page-based features are calculated with all fixation over a given web page.

In order to extract eye gaze features from raw eye-tracking data, a Feature Extraction Tool is designed and developed in Java. It takes eye tracker output, AOI file of the web page, and user demographics file. It has two modules; familiarity and gender. In order to enhance the quality of the data models, the Information Gain method is used to select powerful features and eliminate weak ones. According to information gain method, features which stay under 0.5 threshold are eliminated from the dataset. At the end of feature selection, four features; sequence and page-based Fixations Counts, First Fixated AOI, and Duration of First Fixated AOI are eliminated from datasets for familiarity modeling. Moreover, three features; Sequence based Fixations Counts, First Fixated AOI, and Duration of First Fixated AOI are eliminated from datasets for gender modeling.

In the modeling phase, publicly available eye tracking data from two existing studies are utilized [1] and [2]. There are 79 participants and 6 web pages in total. Each participant assesses each web page twice; for browsing and searching tasks. Those studies intend to create a common eye movement scanpath by using different methods. In this study, the data models are created by two different data mining techniques with 10-fold cross-validation; Logistic Regression and Sequential Minimal Optimization. Because of imbalanced numbers of familiar- unfamiliar users, data models are created for familiarity factor with three web pages; Apple, BBC, and Yahoo. Moreover, for the gender factor, the data models are created for all six web pages. In order to increase the quality of the data models, the datasets are preprocessed by both resampling and SMOTE, synthetic oversampled technique. In modeling, preprocessed datasets cause better data models than a raw dataset.

In the validation phase, the eye tracking study is repeated with new data and test if the data models work or not. The previous studies' procedure is followed to conduct a new eye tracking study and there are 20 participants in total. The data models are validated in two ways. Firstly, the new dataset (20 participants) is added to the existing dataset (79 participants) and 10-fold cross validation is repeated by both Logistic Regression and SMO. Because there are more participants, it is expected that the models become more accurate. Thus, on average, the accuracy of those models is 3% better than the first data models' accuracy. Secondly, the first data model tries to predict the new dataset's dependent variables; familiarity and gender. In predicting, the accuracy values become lower than the expected because data models' accuracies are better. Moreover, on average, the resampled data model's prediction is 40% while the oversampled one is 60% for both factors. However, their precision-recall values are imbalanced which means that prediction is biased. Moreover, the raw data model's prediction is about 50% in average and precision-recall values are quite balanced, but the values are lower than the expected. Therefore, it is seen that searching data models are better than browsing and Logistic Regression trains better than SMO. To conclude, the limitations and future projections of the thesis will be discussed in detail.

6.1 Discussion about Research Questions

This work was conducted in 6 web pages. In totally, the number of participants is 79 for training and 20 for testing. At the beginning of the study, two research questions were generated. Now, they are discussed in this framework.

Research Question 1: Can a familiar user to a web page predicted from the user's eye-tracking data by using data mining techniques?

In the Modeling phase, raw, resampled, and oversampled datasets produced good accuracy values which were supported by good precision, recall, and f-measure values. The datasets are trained by both Logistic Regression and SMO to prevent the problems resourced by algorithms. In these results, resampled datasets on searching trained by Logistic Regression produced better results (*see Section* **??**sec:result1-fam)). Descriptive analyses show that searching datasets of the BBC web page show more significant differences between familiar and unfamiliar participants.

In the Validation phase, new eye-tracking datasets are added to the existing one and they were modeled again firstly. This produced better accuracy values than the Modeling phase. However, in prediction, the accuracy values were not stationary. Data models of both Logistic Regression and SMO cannot produce values as high as the values in the Modeling and Re-Modeling phase. In the baseline analysis, it is seen that a dataset is predicted as either familiar or unfamiliar completely in resampled and oversampled datasets. In the raw datasets, the accuracy is relatively low; but a dataset is predicted as both familiar and unfamiliar with together.

The answer of this research question in the scope of this study is partial yes because the eye-tracking study for validation was conducted in a different place, different times and with different types of eye trackers.

Research Question 2: Can a user predicted as male or female from the user's eye-tracking data by using data mining techniques?

According to the familiarity, modeling, re-modeling, and validation values for gender are lower. Descriptive analyses also show that familiarity has more significant differences on the data. Descriptive analyses show that searching data over Yahoo web page produced more significant differences between females and males. In addition, it can be stated that answer of this research question is partially yes because modeling and re-modeling results are good in spite of lower than the familiarity. The limitations are also valid for gender prediction.

6.2 Limitations

The thesis has three limitations which are mobile eye tracker, different groups of participants, and a small number of participants.

Mobile Eye Tracker: In the validation study, Tobii X2-60 mobile eye tracker has connected to Dell Latitude 7280 12.5" (1366 x 768) laptop because it enables us to conduct the study in different places and find more participants. However, it is different than the eye tracker in modeling study which was Tobii T60 17" (1280 x 1080), built-in eye tracker. The mobile eye tracker sometimes does not record a part of the session and it causes missing recordings in validation. Table 5.2 shows which participant experienced this problem. Moreover, although extracted eye gaze features do not directly show x-y coordinates, distances between fixations may be affected and weaken the validation.

Different Groups of Participants: For the modeling, the previous two eye-tracking studies are utilized and for the validation, a new eye-tracking study is conducted. The first two studies have mostly students because both are conducted in METU NCC and the University of Manchester with mostly 18-24 age group users and they are conducted in English. However, the last eye tracking study is conducted in Ankara with mostly 25-34 age group users and it is conducted in Turkish in order to attract more participants for the validation study. These may affect the validation results, especially the prediction part.

Small Group of Participants: In data mining, the more the amount of data, the more accurate the data models. For instance, in the first phase of the validation study, the new dataset is added to the existing one and 10-fold cross-validation is repeated. Thus, the accuracy of the data models increases 3% more than the modeling one. In order to overcome this problem, resampling and oversampling methods are tried; however, in prediction, they cause bias and cannot predict properly.

In conclusion, this research contributes to the literature with mining eye-tracking data for predicting familiarity and gender of the users. Although the results of modeling and validation are satisfactory, the predictions are not as good as expected. The limitations may be the source of this problem. By repeating the eye tracking study and eliminating these limitations in the future could improve the results and resolves the problems.

6.3 Future Work

This study will be able to be repeated in the future. In this section, the future projections about the thesis are discussed in four phases; repeating the present study, conducting on more web pages, conducting on more user characteristics and a unified model for all web pages.

Firstly, the present study can be repeated with the same inputs and objectives; but, with more participants. In validation, the data models have trained with more participants again and the accuracy values are 3% better than the first data models. This proves that the more the amount of data, the better data models for this study. In this way, the prediction will be enhanced more. If there are more participants, the familiarity may be classified into more ordinal levels rather than the nominal value. Classification to more levels may need to change classifiers such as elastic net, random forest and so on.

Secondly, the eye tracking studies are conducted in six web pages; Apple, AVG, Babylon, BBC, and Yahoo. In this study, their datasets are trained in separate and the data models are not evaluated together. More web pages may be trained and the data models may be combined and generalized for other web pages. Conducting the study on more web pages can verify the results and repeatability.

Thirdly, in this study, the data is classified in terms of familiarity and gender factors. Other user characteristics such as educational background, age, and so on can also be explored. Because of the imbalanced number of the users in this study for educational backgrounds and ages of the participants.

Lastly, in this study, the models are created for each web page separately and the prediction is conducted for a specific web page. As a future work, a unified model can be constructed to predict user's familiarity and gender from their eye movements over any web page. This requires both more participants and more web pages and then a common dataset will be constructed and trained by both algorithms.

REFERENCES

[1] E. S. . Y. Y. . H. S., *Identifying Patterns in Eyetracking Scanpaths in Terms of Visual Elements of Web Pages*, pp. pp 163–180. Springer International Publishing, 2014.

[2] S. Eraslan, Y. Yesilada, and S. Harper, "Scanpath trend analysis on web pages: Clustering eye tracking scanpaths," *ACM Transactions on the Web*, vol. 10, pp. 20:1–20:35, Nov. 2016.

[3] A. Poole and L. J. Ball, "Eye tracking in human-computer interaction and usability research: Current status and future," in *Prospects", Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc*, 2005.

[4] B. Turgay and T. Hakan, *Tüketicileri (kullanıcıları) ve ürün kullanımlarını analiz etmek için göz izleme yönteminin kullanılması*. 01 2014.

[5] H. H. Greene and K. Rayner, "Eye movements and familiarity effects in visual search," *Vision Research*, vol. 41, no. 27, pp. 3763 – 3773, 2001.

[6] F. Marchal, S. Castagnos, and A. Boyer, "Tell me what you see, i will tell you what you remember," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, UMAP '16, pp. 293–294, ACM, 2016.

[7] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman, "The determinants of web page viewing behavior: An eye-tracking study," in *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ETRA '04, pp. 147–154, ACM, 2004.

[8] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *Proceedings of the 12th Web for All Conference*, W4A '15, (New York, NY, USA), pp. 16:1–16:8, ACM, 2015.

[9] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting autism based on eye-tracking data from web searching tasks," in *Proceedings of the Internet of Accessible Things*, W4A '18, (New York, NY, USA), pp. 16:1–16:10, ACM, 2018.

[10] L. G. . M. N., "Cognitive determinants of fixation location during picture viewing," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 4, pp. 565–572, 1978.

[11] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychological bulletin*, vol. 124, pp. 372–422, 12 1998.

[12] S. Josephson and M. E. Holmes, "Visual attention to repeated internet images: Testing the scanpath theory on the world wide web," in *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, (New York, NY, USA), pp. 43–49, ACM, 2002.

[13] Y. Yesilada, C. Jay, R. Stevens, and S. Harper, "Validating the use and role of visual elements of web pages in navigation with an eye-tracking study," in *The 17th International Conference on World Wide Web. WWW '08*, pp. 11–20, 2008.

[14] J. Salojärvi, I. Kojo, J. Simola, and S. Kaski, "Can relevance be inferred from eye movements in information retrieval?," 2003.

[15] D. Santella, "Robust clustering of eye movement recordings for quantification of visual interest," in *Eye Tracking Research and Applications (ETRA) Symposium 2004*, 2004.

[16] D. Bondareva, C. Conati, R. Feyzi-Behnagh, J. M. Harley, R. Azevedo, and F. Bouchet, *Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning*, pp. 229–238. Springer Berlin, 2013.

[17] B. Steichen, M. M. A. Wu, D. Toker, C. Conati, and G. Carenini, *Te,Te,Hi,Hi: Eye Gaze Sequence Analysis for Informing User-Adaptive Information Visualizations*, pp. 183–194. Cham: Springer International Publishing, 2014.

[18] D. Toker, C. Conati, B. Steichen, and G. Carenini, "Individual user characteristics and information visualization: Connecting the dots through eye tracking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 295–304, ACM, 2013.

[19] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *International Journal of Industrial Ergonomics*, vol. 24, pp. 631–645, Oct. 1999.

[20] S. Eraslan and Y. Yesilada, "Patterns in eyetracking scanpaths and the affecting factors," *Journal of Web Engineering*, vol. 14, pp. 363–385, November 2015.

[21] R. J. Jacob and K. S. Karn, "Eye tracking in Human-Computer interaction and usability research: Ready to deliver the promises," *The Mind's eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pp. 573–603, 2003.

[22] A. Poole, L. J. Ball, and P. Phillips, *In Search of Salience: A Response-time and Eye-movement Analysis of Bookmark Recognition*, pp. 363–378. London: Springer London, 2005.

[23] L. Cowen, L. J. Ball, and J. Delin, *An Eye Movement Analysis of Web Page Usability*, pp. 317–335. Springer London, 2002.

[24] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky, "Eye tracking in web search tasks: Design implications," in *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, (New York, NY, USA), pp. 51–58, ACM, 2002.

[25] J. McDonald, *Handbook of Biological Statistics (3rd ed).* Sparky House Publishing, 2014.

[26] D. R. Cox, "The regression analysis of binary sequences (with discussion)," *J Roy Stat Soc B*, vol. 20, pp. 215–242, 1958.

[27] V. Vapnik, *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics).* Berlin, Heidelberg: Springer-Verlag, 1982.

[28] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[29] S. Ding, H. Zhu, W. Jia, and C. Su, "A survey on feature extraction for pattern recognition," *Artificial Intelligence Review*, vol. 37, pp. 169–180, Mar 2012.

[30] Y. he, B. Gao, A. Sophian, and R. Yang, *Magnetic Sensor Based Pulsed Eddy Current for Defect Detection and Characterization*, pp. 7–35. 12 2017.

[31] D. Cai, S. Yu, J.-R. Wen, and W.-Y. Ma, "Vips: a vision-based page segmentation algorithm," Tech. Rep. MSR-TR-2003-79, November 2003.

[32] E. Michailidou, "Vicram: Visual complexity rankings and accessibility metrics," *SIGACCESS Access. Comput.*, pp. 24–27, Sept. 2006.

[33] M. E. Akpinar and Y. Yesilada, "Vision based page segmentation algorithm: Extended and perceived success," in *The ICWE 2013 International Workshops on Current Trends in Web Engineering - Volume 8295*, pp. 238–252, Springer-Verlag, 2013.

[34] S. le Cessie and J. van Houwelingen, "Ridge estimators in logistic regression," *Applied Statistics*, vol. 41, no. 1, pp. 191–201, 1992.

[35] E. Frank, M. Hall, H. Geoffrey, R. Kirkby, B. PfahringerIan, and H. W. Trigg, "Weka-a machine learning workbench for data mining," *Data Mining and Knowledge Discovery Handbook*, 2010.

[36] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, pp. 321–357, June 2002.

[37] G. B. Demisse, T. Tadesse, and Y. Bayissa, "Data mining attribute selection approach for drought modeling: A case study for greater horn of africa," *CoRR*, vol. abs/1708.05072, 2017.

[38] S. McLeod, "Sampling methods." Retrieved from https://www.simplypsychology.org/sampling.html, 2014.

[39] J. Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15*. Milton Keynes, UK, USA: Open University Press, 3rd ed., 2007.

[40] G. Hole, "Research skills," 2019.

[41] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[42] P. Brusilovsky, *From Adaptive Hypermedia to the Adaptive Web*, pp. 21–24. Wiesbaden: Vieweg+Teubner Verlag, 2003.

Appendix A

TOBII EYE TRACKER OUTPUT SAMPLE

Figure A.1 Tobii Eye Tracker Output Sample

FixationIndex	TimaStamp	FixationDuration	MappedFixationPointX	MappedFixationPointY	StimuliName
1	33	550	0	0	No Media
2	583	218	0	0	No Media
3	801	250	283	545	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
4	1051	100	240	629	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
5	1151	200	451	264	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
6	1351	400	586	276	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
7	1751	483	553	175	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
8	2234	350	639	324	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
9	2584	234	357	90	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
10	2818	150	313	101	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
11	2968	300	509	69	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
12	3268	383	671	356	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
13	3651	200	717	373	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
14	3851	167	632	533	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
15	4018	483	695	563	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
16	4501	200	704	397	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
17	4701	233	700	335	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
18	4934	217	620	320	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
19	5151	250	356	500	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
20	5401	167	324	596	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
21	5568	149	302	581	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
22	5717	652	729	377	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
23	6369	417	436	153	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
24	6786	533	339	205	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
25	7319	217	629	277	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
26	7536	83	665	285	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
27	7619	883	724	393	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
28	8502	584	608	316	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
29	9086	433	320	215	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)
30	9519	284	354	214	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/ (CRC)

Appendix B

AREA OF INTEREST FILE SAMPLE

Figure B.1 Area of Interest File Sample

```
VB.1.1.1 998 179 11 47 A
VB.1.1.2 185 813 11 47 B
VB.1.2.1.1.1 207 453 178 212 C
VB.1.2.1.1.2.1 258 157 433 115 D
VB.1.2.1.1.2.2 418 157 433 115 E
VB.1.2.1.2 661 513 70 621 F
VB.1.2.2.1 183 250 700 150 G
VB.1.2.2.2 438 250 700 150 H
VB.1.2.2.3 688 250 700 150 I
VB.1.2.2.4 938 250 700 150 J
```

Appendix C

USER DEMOGRAPHICS FILE SAMPLE

Figure C.1 User Demographics File Sample (Validation Study Participants)

ID	Gender	Age Group	Education Level	Godaddy	Apple	AVG	Yahoo	Babylon	BBC
1	Female	18-24	High School	5	5	5	1	5	4
2	Female	25-34	Undergraduate	4	1	5	1	5	5
3	Male	35-54	Graduate	5	4	5	1	5	3
4	Female	35-54	Middle School	5	5	5	5	5	5
5	Female	25-34	High School	4	4	4	1	3	3
6	Male	25-34	Graduate	5	3	5	4	5	2
7	Male	25-34	Undergraduate	5	3	3	1	4	2
8	Male	25-34	Undergraduate	5	1	3	1	1	1
9	Male	35-54	Undergraduate	4	5	5	5	5	5
10	Male	25-34	Undergraduate	5	3	4	1	5	1
11	Male	35-54	Undergraduate	5	2	4	3	5	3
12	Female	35-54	Graduate	5	1	2	3	4	5
13	Male	35-54	High School	5	4	3	4	5	3
14	Female	35-54	Undergraduate	5	4	5	3	5	4
15	Female	25-34	High School	5	4	5	4	5	4
16	Male	25-34	High School	2	1	3	1	5	5
17	Male	35-54	High School	5	3	5	4	5	5
18	Male	25-34	Undergraduate	4	4	5	4	4	1
19	Female	18-24	High School	5	2	5	2	1	2
20	Male	25-34	High School	4	1	4	4	5	1

Appendix D

INFORMATION SHEET OF THE PREVIOUS STUDIES

Participant Information Sheet

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Who will conduct the research? Sukru Eraslan

Title of the research. Understanding Eye Tracking Data for Re-Engineering Web Pages

Why have I been chosen? I am inviting anyone who is computer literate and between the ages of 18 and 35 to take part in the evaluation if they want to.

What would I be asked to do if I take part? You will be asked to fill a short questionnaire about your demographic information and Web experience. Next, you will be required to complete two simple tasks on three Web pages and just scan other three Web pages while your eye movements are tracked. The pages will be shown to you and the investigator will ask you to complete a task, which can be a particular task or just scanning, on the pages. At the end, your opinions about the Web pages will be asked.

What happens to the data collected? Electronic data will be stored securely on a computer. Written information will be stored in a locked drawer. The information will be used in preparation of my dissertation.

How is confidentiality maintained? Data will be made anonymous (names and any other information that may identify an individual will not be included), so no one will be able to recognize who the data belongs to.

Will I be paid for participating in the research? You will not be paid for participating in the research.

What is the duration of the research? The study will take less than one hour to complete.

Where will the research be conducted? Room: SZ-06, Academic S Building Middle East Technical University Northern Cyprus Campus Academic

What if I change my mind? It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself.

Will the outcomes of the research be published? The outcomes of the research will be published in my thesis and conference proceedings and journal articles.

Contact for further information. For further information, please contact either myself or my supervisor (details above).

What if something goes wrong? If a participant wants to make a formal complaint about the conduct of the research they should contact the Head of the Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL

Appendix E

CONSENT FORM OF THE PREVIOUS STUDIES

Figure E.1 Consent Form of the Previous Studies

MANCHESTER 1824

The University of Manchester

School of Computer Science

Understanding Eye Tracking Data for Re-engineering Web Pages

CONSENT FORM

If you are happy to participate please complete and sign the consent form below

		Please Initial Box
1	I confirm that I have read the attached information sheet on the above project and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason.	
3	I understand that the session will be audio recorded and an eye-tracker will be used.	
4	I agree to the use of anonymous quotes.	

I agree to take part in the above project.

Name of participant	Date	Signature

Name of person taking consent	Date	Signature

Appendix F

QUESTIONNAIRE OF THE PREVIOUS STUDIES

Eye Tracking Study - Questionnaire

1. What is your gender?

- Female
- Male

2. What is your age?

- Under 18
- 18-24
- 25-34
- 35-54
- 55+

3. How often do you use the Web?

- Daily
- Weekly
- Monthly
- Less than once a month
- Never

4. Highest level of education you have completed:

- Grade/Primary School

- High/Secondary School
- Associates Degree
- Bachelors Degree
- Masters Degree
- Doctorate
- Other

5. How often do you visit the web site? (Your answer should be 1,2,3,4 or 5)

- 1 : Daily
- 2 : Weekly
- 3 : Mountly
- 4 : Less than once a month
- 5 : Never

- Go Daddy<www.godaddy.com/> ———-
- Apple <http/www.apple.com/> ———-
- AVG <www.avg.com/us-en/index.html> ———-
- Yahoo! <http/www.yahoo.com/> ———-
- babylon <http/www.babylon.com/> ———-
- BBC <http/www.bbc.co.uk/> ———-

Appendix G

INFORMATION SHEET OF THE VALIDATION STUDY

Bilgilendirme Metni

Araştırma projemize katılmaya sizi davet ediyoruz. Fakat, projeye katılmadan önce sizi projenin amacı ve kapsamı konusunda bilgilendirmek istiyoruz.

Lütfen aşağıdaki bilgilendirmeyi dikkatle okuyunuz. Kafanıza takılan herhangi bir soru olması halinde, lütfen bizimle iletişime geçiniz.
MELİH ÖDER

Araştırmanın Adı Nedir? Göz Takibi Verileri Madenciliği ile Kullanıcıları ve Kullanma Alışkanlıklarını Karaketirize Etme

Neden ben seçildim? Çalışmamıza bilgisayar okuryazarlığı olan 18 – 55 yaş aralığında katılımcıları davet ediyoruz.

Katılırsam ne yapacağım? Öncelikle demografik bilgilerinizi (yaş, cinsiyet, vb.) ve internet deneyiminizi paylaşmanızı isteyeceğimiz kısa bir anket dolduracaksınız. Bu bilgileriniz kesinlikle 3. kişilerle paylaşılmayacaktır. Anonim bilgi olarak çalışma içerisinde kullanılacaktır.Sonrasında 6 ayrı web sayfasını gözlerinizle bir süre taramanızı isteyeceğiz. Ardından her bir web sayfası üzerinde 2'şer basit görevi tamamlamanızı bekleyeceğiz. Bu arada göz izleme verileriniz kayıt altında olacaktır. Görevlerin sonunda deneyim ile alakalı düşüncelerinizi bizimle paylaşmanızı isteyeceğiz.

Kayıt edilen verilerimiz nasıl saklanacak? Veriler isim kullanılmaksızın, kodlarla isimlendirilen dosyalar içinde güvenli bir bilgisayarda (kişisel bilgisayarım), şifrelenmiş bir klasör içinde barındırılacaktır. Dosya isimleri yalnızca benim bildiğim bir kurala göre kodlu olarak verilecektir. Kesinlikle kişiyi açık eden bir formatta olmayacaktır.

Çalışma ne kadar sürecek? Yaklaşık 15 dk. sürecek bir çalışmadır.

Araştırma nerede yapılacak? ODTÜ Bilgi İşlem Dairesi (Computer Center – Bilgisayar Müh. yanı), 1. Katta yer alan İnsan-Bilgisayar Etkileşimi Laboratuvarında yapılacaktır.

Araştırma verileri nasıl kullanılacak? Araştırma verilerini ODTÜ Bilişim Sistemleri A.B.D'sinde yürüttüğüm yüksek lisans tezim kapsamında kullanacağım.

Appendix H

CONSENT FORM OF THEVALIDATION STUDY

Figure H.1 Consent Form of the Validation Study

ORTA DOĞU TEKNİK ÜNİVERSİTESİ

ENFORMATİK ENSTİTÜSÜ / BİLİŞİM SİSTEMLERİ ANABİLİM DALI

GÖZ TAKİBİ VERİLERİ MADENCİLİĞİ İLE KULLANICILARI VE KULLANMA ALIŞKANLIKLARINI KARAKTERİZE ETME

Katılım Formu

Eğer bu çalışmaya katılmaya istekliyseniz, aşağıdaki katılım formunu doldurup, imzalamalısınız.

	Evet	Hayır
1. Bu formla birlikte verilen proje ile alakalı bilgilendirme metnini okudum. Proje ile alakalı soru sormam ve bilgi almam sağlandı. Sorularım tatmin edici bir şekilde yanıtlandı.	☐	☐
2. Çalışmaya gönüllü olarak katılıyorum ve istediğim zaman herhangi bir sebep belirtmeksizin çalışmadan çekilebileceğimi anladım.	☐	☐
3. Çalışma boyunca ses kaydı yapılacağını ve göz hareketlerimin kaydedileceğini anladım.	☐	☐
4. Kullandığım cümlelerin anonim olarak çalışma içinde kullanılabileceğini anladım.	☐	☐

Lütfen "Yukarıdaki çalışmaya gönüllü olarak katılıyorum." İfadesini el yazınızla yazınız.

Tarih / İmza

Appendix I

QUESTIONNAIRE OF THE VALIDATION STUDY

Göz İzleme Çalışması - Anket

1. Cinsiyetiniz nedir?

- Kadın
- Erkek

2. Yaş grubunuzu belirtiniz?

- 18 yaşından küçük
- 18-24
- 25-34
- 35-54
- 55 yaşından büyük

3. İnternet kullanma sıklığınızı belirtiniz.?

- Hergün
- Haftada birkaç kez
- Ayda birkaç kez
- Hiç

4. Eğitim seviyenizi belirtiniz:

- İlköğretim
- Orta Öğretim

- Ön Lisans
- Lisans
- Yüksek Lisans
- Doktora

5. Aşağıdaki web sayfalarını ne sıklıkla kullandığınızı skalayı kullanarak belirtiniz? (1 - 5 aralığında olmalıdır)

- 1 : Günlük
- 2 : Haftalık
- 3 : Aylık
- 4 : Ayda bir defadan daha az
- 5 : Hiç

- Go Daddy<www.godaddy.com/> ———-
- Apple <http/www.apple.com/> ———-
- AVG <www.avg.com/us-en/index.html> ———-
- Yahoo! <http/www.yahoo.com/> ———-
- babylon <http/www.babylon.com/> ———-
- BBC <http/www.bbc.co.uk/> ———-

Appendix J

WEB PAGES AND AOIS

Figure J.1 Apple Web Page and AOIs

Figure J.2 AVG Web Page and AOIs

Figure J.3 Babylon Web Page and AOIs

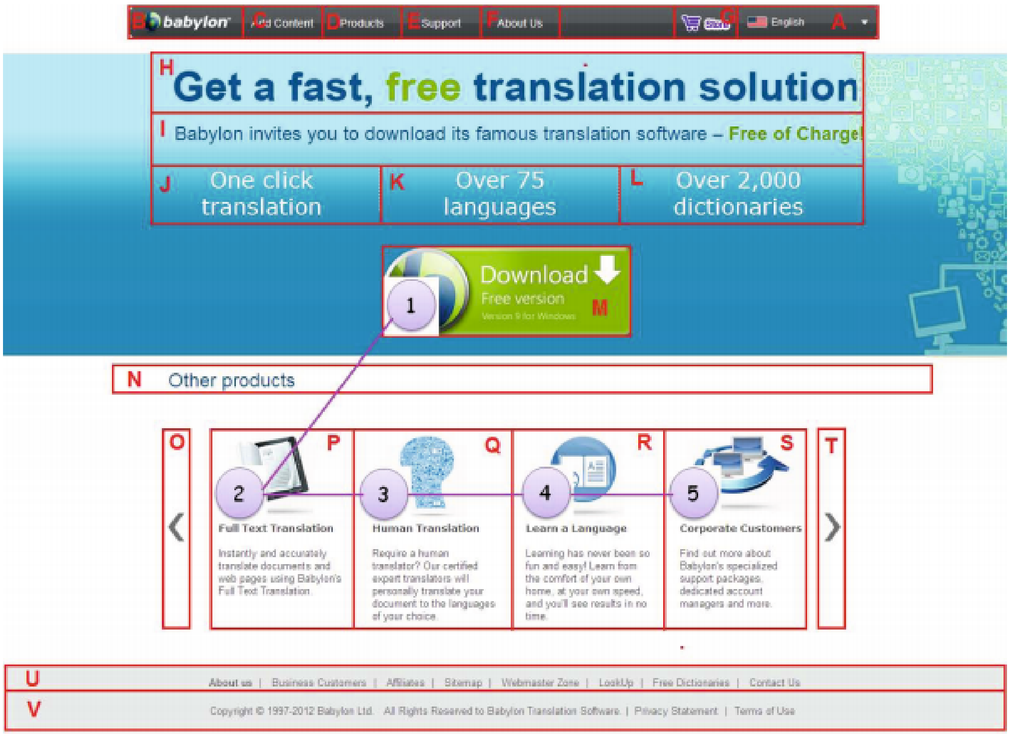


Figure J.4 AVG Web Page and AOIs

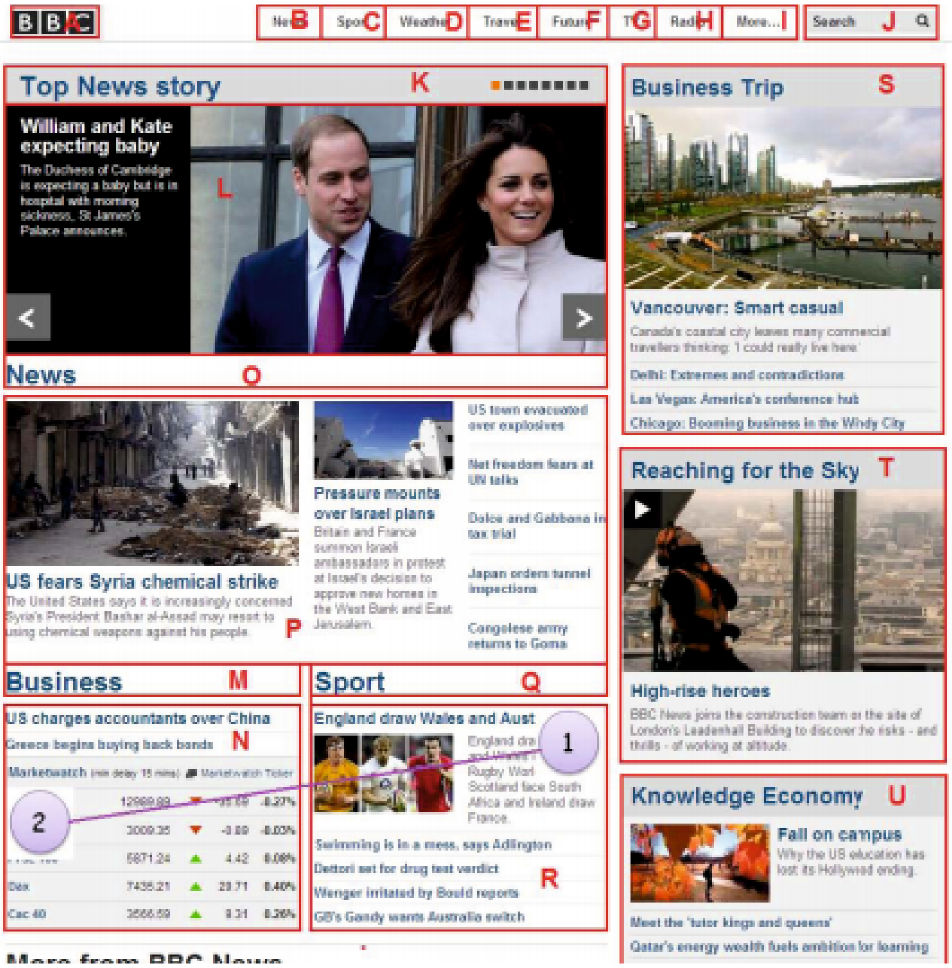


Figure J.5 Godaddy Web Page and AOIs

Figure J.6 Yahoo Web Page and AOIs

Appendix K

BASELINE ANALYSIS TABLES OF THE FAMILIARITY PREDICTION

Table K.1: Apple - Browsing - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	1	0	1	1	1	0	0
2	0	0	1	1	1	0	0
3	1	0	1	1	1	0	0
4	1	0	1	1	1	0	0
5	1	1	1	1	1	0	0
6	0	0	1	1	1	0	0
7	0	0	1	1	1	0	0
8	0	0	1	1	1	0	0
9	1	0	1	1	1	0	0
10	0	0	1	1	1	0	0
11	0	0	1	1	1	0	0
12	0	1	1	1	1	0	0
13	1	0	1	1	1	0	0
14	1	1	1	1	1	0	0
15	1	0	1	1	1	0	0

Table K.2: Apple - Searching - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	1	0	1	1	1	0	0
2	0	0	1	1	1	0	0
3	1	0	1	1	1	0	0
4	1	0	1	1	1	0	0
5	1	0	1	1	1	0	0
6	0	0	1	1	1	0	0
7	0	0	1	1	1	0	0
8	0	0	1	1	1	0	0
9	1	0	1	1	1	0	0
10	0	0	1	1	1	0	0
11	0	0	1	1	1	0	0
12	0	1	1	1	1	0	0
13	1	0	1	1	1	0	0
14	1	0	1	1	1	0	0
15	1	0	1	1	1	0	0
16	0	0	1	1	1	0	0

Table K.3: BBC - Browsing - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	1	1	0	1	1	1	1
2	1	1	0	1	1	1	1
3	0	1	0	1	1	1	1
4	1	0	0	1	1	1	1
5	0	1	0	1	1	1	1
6	0	1	0	1	1	1	1
7	0	1	0	1	1	1	1
8	0	1	0	1	1	1	1
9	1	1	0	1	1	1	1
10	0	1	0	1	1	1	1
11	0	1	0	1	1	1	1
12	1	1	0	1	1	1	1
13	0	1	0	1	1	1	1
14	1	1	0	1	1	1	1
15	1	1	0	1	1	1	1
16	1	1	0	1	1	1	1
17	1	1	0	1	1	1	1

Table K.4: BBC - Searching - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	1	1	0	1	1	1	1
2	1	1	0	1	1	1	1
3	0	1	0	1	1	1	1
4	1	1	0	1	1	1	1
5	0	1	0	1	1	1	1
6	0	1	0	1	1	1	1
7	0	1	0	1	1	1	1
8	0	1	0	1	1	1	1
9	1	0	0	1	1	1	1
10	0	1	0	1	1	1	1

Table K.5: Yahoo - Browsing - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	0	1	0	0	0	0
2	0	0	1	0	0	0	0
3	0	1	1	0	0	0	0
4	1	0	1	0	0	0	0
5	0	0	1	0	0	0	0
6	1	0	1	0	0	0	0
7	0	0	1	0	0	0	0
8	0	0	1	0	0	0	0
9	1	0	1	0	0	0	0
10	0	0	1	0	0	0	0
11	0	0	1	0	0	0	0
12	0	0	1	0	0	0	0
13	1	0	1	0	0	0	0
14	0	1	1	0	0	0	0
15	1	0	1	0	0	0	0
16	0	0	1	0	0	0	0
17	1	0	1	0	0	0	0
18	1	0	0	0	0	0	0

Table K.6: Yahoo - Searching - Prediction Table (0: Familiar, 1: Unfamiliar)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	1	1	1	1	0	0
2	0	1	1	1	1	0	0
3	0	1	1	1	1	0	0
4	1	1	1	1	1	0	0
5	0	1	1	1	1	0	0
6	1	1	1	1	1	0	0
7	0	1	1	1	1	0	0
8	0	1	1	1	1	0	0
9	1	1	1	0	1	0	0
10	0	1	1	1	1	0	0
11	0	1	1	1	1	0	0
12	0	1	1	1	1	0	0
13	1	1	1	1	1	0	0
14	0	1	1	1	1	0	0
15	1	1	1	1	1	0	0
16	0	1	1	1	1	0	0
17	1	1	1	1	1	0	0
18	1	1	1	1	1	0	0
19	0	1	1	1	1	0	0

Appendix L

BASELINE ANALYSIS TABLES OF THE GENDER PREDICTION

Table L.1: Apple - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	0	0	0	0	1	1
2	0	1	1	0	0	1	1
3	1	0	1	0	0	1	1
4	0	0	1	0	0	1	1
5	0	1	1	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	1	0	0	1	1
8	1	0	0	0	0	1	1
9	1	0	1	0	0	1	1
10	1	0	1	0	0	1	1
11	1	0	1	0	0	1	1
12	0	0	0	0	0	1	1
13	1	0	1	0	0	1	1
14	0	0	1	0	0	1	1
15	0	0	0	0	0	1	1

Table L.2: Apple - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	1	1	0	0	1	1
2	0	1	1	0	0	1	1
3	1	1	0	0	0	1	1
4	0	1	1	0	0	1	1
5	1	1	1	0	0	1	1
6	1	1	1	0	0	1	1
7	1	1	1	0	0	1	1
8	1	1	1	0	0	1	1
9	1	1	0	0	0	1	1
10	0	1	1	0	0	1	1
11	1	1	1	0	0	1	1
12	0	1	1	0	0	1	1
13	1	1	1	0	0	1	1
14	1	1	1	0	0	1	1
15	0	1	1	0	0	1	1
16	1	0	0	0	0	1	1

Table L.3: AVG - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	1	0	0	0	1	1
2	0	1	0	0	0	1	1
3	1	0	0	0	0	1	1
4	0	1	0	0	0	1	1
5	0	1	0	0	0	1	1
6	1	1	0	0	0	1	1
7	1	1	0	0	0	1	1
8	1	1	0	0	0	1	1
9	0	1	0	0	0	1	1
10	1	0	0	0	0	1	1
11	0	0	0	0	0	1	1
12	1	0	0	0	0	1	1
13	1	1	0	0	0	1	1
14	1	1	0	0	0	1	1
15	0	1	0	0	0	1	1
16	1	1	0	0	0	1	1

Table L.4: AVG - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	1	1	0	0	1	1
2	0	0	0	0	0	1	1
3	1	0	0	0	0	1	1
4	0	0	0	0	0	1	1
5	0	0	0	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	0	0	0	1	1
8	1	1	1	0	0	1	1
9	1	0	0	0	0	1	1
10	1	0	0	0	0	1	1
11	1	0	0	0	0	1	1
12	0	0	0	0	0	1	1
13	1	0	0	0	0	1	1
14	0	0	0	0	0	1	1
15	0	0	0	0	0	1	1
16	1	1	0	0	0	1	1
17	1	0	0	0	0	1	1
18	1	0	0	0	0	1	1
19	0	0	0	0	0	1	1
20	1	0	0	0	0	1	1

Table L.5: Babylon - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	0	0	0	0	1	1
2	0	1	1	0	0	1	1
3	1	0	0	0	0	1	1
4	0	0	0	0	0	1	1
5	0	1	1	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	0	0	0	1	1
8	1	1	1	0	0	1	1
9	1	0	0	0	0	1	1
10	1	0	0	0	0	1	1
11	1	0	0	0	0	1	1
12	0	1	1	0	0	1	1
13	1	1	1	0	0	1	1
14	0	0	0	0	0	1	1
15	0	1	1	0	0	1	1
16	1	1	1	0	0	1	1
17	1	0	0	0	0	1	1
18	1	1	1	0	0	1	1
19	0	0	1	0	0	1	1
20	1	0	0	0	0	1	1

Table L.6: Babylon - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	0	0	0	0	1	1
2	0	0	0	0	0	1	1
3	1	0	0	0	0	1	1
4	0	0	0	0	0	1	1
5	0	0	0	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	0	0	0	1	1
8	1	0	0	0	0	1	1
9	1	0	0	0	0	1	1
10	1	0	0	0	0	1	1
11	1	0	0	0	0	1	1
12	0	1	0	0	0	1	1
13	1	0	0	0	0	1	1
14	0	0	0	0	0	1	1
15	0	0	0	0	0	1	1
16	1	0	0	0	0	1	1
17	1	0	0	0	0	1	1
18	1	0	0	0	0	1	1
19	0	0	0	0	0	1	1
20	1	0	0	0	0	1	1

Table L.7: BBC - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	0	0	0	0	0	1
2	0	0	0	0	0	0	1
3	1	1	0	0	0	0	1
4	0	0	0	0	0	0	1
5	0	0	0	0	0	0	1
6	1	1	0	0	0	0	1
7	1	1	0	0	0	0	1
8	1	1	0	0	0	0	1
9	1	1	0	0	0	0	1
10	0	0	0	0	0	0	1
11	1	1	0	0	0	0	1
12	0	0	1	1	0	0	1
13	1	1	1	0	0	0	1
14	1	1	0	0	0	0	1
15	1	1	1	1	0	0	1
16	0	0	0	0	0	0	1
17	1	1	1	1	0	0	1

Table L.8: BBC - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	1	1	0	0	1	1
2	1	1	1	0	0	1	1
3	0	1	1	0	0	1	1
4	0	0	0	0	0	1	1
5	1	1	1	0	0	1	1
6	1	0	0	0	0	1	1
7	0	1	1	0	0	1	1
8	1	1	1	0	0	1	1
9	1	0	0	0	0	1	1
10	1	1	1	0	0	1	1

Table L.9: GoDaddy - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	1	1	0	0	1	1
2	0	0	0	0	0	1	1
3	1	1	1	0	0	1	1
4	0	1	1	0	0	1	1
5	0	1	1	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	0	0	0	1	1
8	1	1	1	0	0	1	1
9	1	1	1	0	0	1	1
10	1	1	1	0	0	1	1
11	1	1	1	0	0	1	1
12	0	1	1	0	0	1	1
13	1	0	0	0	0	1	1
14	0	0	0	0	0	1	1
15	0	1	1	0	0	1	1
16	1	1	1	0	0	1	1
17	1	0	0	0	0	1	1
18	1	0	0	0	0	1	1
19	0	0	0	0	0	1	1
20	1	1	1	0	0	1	1

Table L.10: GoDaddy - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	Actual Value	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO	Predicted by Logistic Regression	Predicted by SMO
1	0	1	1	1	1	1	1
2	0	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	0	1	1	1	1	1	1
5	0	1	1	1	1	1	1
6	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1
11	0	1	1	1	1	1	1
12	1	1	1	1	1	1	1
13	0	1	1	1	1	1	1
14	0	1	1	1	1	1	1
15	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1
17	0	1	1	1	1	1	1
18	1	1	1	1	1	1	1

Table L.11: Yahoo - Browsing - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	1	1	0	0	1	1
2	0	0	0	0	0	1	1
3	1	0	0	0	0	1	1
4	0	1	1	0	0	1	1
5	1	1	0	0	0	1	1
6	1	0	0	0	0	1	1
7	1	0	0	0	0	1	1
8	1	0	0	0	0	1	1
9	1	0	0	0	0	1	1
10	1	0	0	0	0	1	1
11	0	0	0	0	0	1	1
12	1	1	1	0	0	1	1
13	0	1	1	0	0	1	1
14	1	1	1	0	0	1	1
15	1	1	0	0	0	1	1
16	1	0	0	0	0	1	1
17	0	0	0	0	0	1	1
18	1	0	0	0	0	1	1

Table L.12: Yahoo - Searching - Prediction Table (0: Female, 1: Male)

		Raw Dataset		Resampled Dataset		Oversampled Dataset	
ID	**Actual Value**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**	**Predicted by Logistic Regression**	**Predicted by SMO**
1	0	1	1	0	0	1	1
2	0	1	1	0	0	1	1
3	1	1	1	0	0	1	1
4	0	1	1	0	0	1	1
5	0	1	1	0	0	1	1
6	1	1	1	0	0	1	1
7	1	1	1	0	0	1	1
8	1	1	1	0	0	1	1
9	1	1	1	0	0	1	1
10	1	1	1	0	0	1	1
11	1	1	1	0	0	1	1
12	0	1	1	0	0	1	1
13	1	1	1	0	0	1	1
14	0	1	1	0	0	1	1
15	1	1	1	0	0	1	1
16	1	1	1	0	0	1	1
17	1	1	1	0	0	1	1
18	0	1	1	0	0	1	1
19	1	1	1	0	0	1	1

Appendix M

DESCRIPTIVE ANALYSIS OF TRAINING DATA - FAMILIARITY

Table M.1: Descriptive Analysis of Familiarity with 79 Participants

		Mean of SB Fixation Duration (ms)	Mean of PB Fixation Duration (ms)	Scanpath Length (num. of)	Mean of PB Fixation Counts (num. of)	Mean of PB Fixation Distances (pxl)	Mean of PB Path Angles (degree)
Apple - Browsing	Familiar	**282** (sd. 71)	**325** (sd. 74)	**19** (sd. 9)	**86** (sd. 15)	**201** (sd. 33)	**6.73** (sd. 12)
	Unfamiliar	**297** (sd. 74)	**314** (sd. 53)	**20** (sd.8)	**86** (sd. 15)	**208** (sd. 44)	**4.86** (sd. 14)
BBC - Browsing	Familiar	**322** (sd. 64)	**314** (sd. 53)	**61** (sd. 13)	**90** (sd. 13)	**181** (sd. 34)	**0.24** (sd. 11)
	Unfamiliar	**311** (sd. 46)	**310** (sd. 41)	**60** (sd. 13)	**90** (sd. 11)	**181** (sd. 28)	**-1.76** (sd. 11)
Yahoo - Browsing	Familiar	**303** (sd. 77)	**339** (sd. 75)	**15** (sd. 7)	**85** (sd. 16)	**175** (sd. 31)	**1.75** (sd. 13)
	Unfamiliar	**290** (sd. 73)	**342** (sd. 66)	**19** (sd. 10)	**84** (sd. 15)	**173** (sd. 23)	**0.68** (sd. 12)
Apple - Searching	Familiar	**389** (sd. 78)	**355** (sd. 38)	**26** (sd. 10)	**78** (sd. 34)	**207** (sd. 30)	**3.24** (sd. 13)
	Unfamiliar	**392** (sd. 128)	**336** (sd. 56)	**31** (sd. 13)	**81** (sd. 35)	**208** (sd. 25)	**0.15** (sd. 14)
BBC - Searching	Familiar	**269** (sd. 48)	**325** (sd. 52)	**39** (sd. 11)	**40** (sd. 22)	**83** (sd. 31)	**-2.82** (sd. 11)
	Unfamiliar	**283** (sd. 42)	**340** (sd. 49)	**41** (sd. 25)	**83** (sd. 41)	**179** (sd. 22)	**-1.26** (sd. 13.5)
Yahoo - Searching	Familiar	**294** (sd. 64)	**436** (sd. 89)	**26** (sd. 26)	**114** (sd. 60)	**162** (sd. 24)	**-0.11** (sd. 12)
	Unfamiliar	**287** (sd. 74)	**429** (sd. 72)	**26** (sd. 23)	**108** (sd. 46)	**159** (sd. 24)	**-3.48** (sd. 11)

Appendix N

DESCRIPTIVE ANALYSIS OF TRAINING DATA - GENDER

Table N.1: Descriptive Analysis of Gender with 79 Participants

		Mean of SB Fixation Duration (ms)	Mean of PB Fixation Duration (ms)	Scanpath Length (num. of)	Mean of PB Fixation Counts (num. of)	Mean of PB Fixation Distances (pxl)	Mean of PB Path Angles (degree)
Apple - Browsing	Female	**297** (sd. 68)	**315** (sd. 49)	**20** (sd. 9)	**86** (sd. 17)	**212** (sd. 50)	**8.87** (sd. 15)
	Male	**287** (sd. 80)	**321** (sd. 69)	**20** (sd. 8)	**86** (sd. 13)	**200** (sd. 29)	**1.89** (sd. 11)
AVG - Browsing	Female	**316** (sd. 66)	**335** (sd. 68)	**33** (sd. 12)	**86** (sd. 15)	193 (sd. 35)	**-1.54** (sd. 10)
	Male	**288** (sd. 62)	**304** (sd. 48)	**37** (sd. 14)	**91** (sd. 13)	182 (sd. 28)	**-0.30** (sd. 12)
Babylon - Browsing	Female	**291** (sd. 61)	**324** (sd. 70)	**44** (sd. 17)	**88** (sd. 16)	**179** (sd. 32)	**1.44** (sd. 11)
	Male	**294** (sd. 49)	**311** (sd. 45)	**48** (sd. 16)	**91** (sd. 13)	**181** (sd. 30)	**3.57** (sd. 11)
BBC - Browsing	Female	**323** (sd. 52)	**318** (sd. 51)	**61** (sd. 13)	**89** (sd. 11)	**186** (sd. 32)	**-2.68** (sd. 10)
	Male	**309** (sd. 59)	**307** (sd. 50)	**60** (sd. 13)	**91** (sd. 13)	**176** (sd. 30)	**1.30** (sd. 12)
GoDaddy - Browsing	Female	**312** (sd. 64)	**318** (sd. 48)	**53** (sd. 15)	**87** (sd. 13)	**191** (sd. 28)	**0.42** (sd. 13)
	Male	**304** (sd. 78)	**304** (sd. 58)	**54** (sd. 20)	**89** (sd. 15)	**187** (sd. 30)	**1.29** (sd. 11)
Yahoo - Browsing	Female	**310** (sd. 77)	**342** (sd. 64)	**17** (sd. 10)	**83** (sd. 15)	**178** (sd. 26)	**1.01** (sd. 14)
	Male	**280** (sd. 67)	**338** (sd. 74)	**18** (sd. 9)	**85** (sd. 14)	**170** (sd. 27)	**1.07** (sd. 11)
Apple - Searching	Female	**394** (sd. 76)	**344** (sd. 46)	**29** (sd. 12)	**81** (sd. 34)	**210** (sd. 26)	**1.40** (sd. 14)
	Male	**390** (sd. 145)	**340** (sd. 57)	**29** (sd. 12)	**80** (sd. 34)	**206** (sd. 25)	**0.21** (sd. 13)
AVG - Searching	Female	**328** (sd. 89)	**318** (sd. 42)	**32** (sd. 18)	**86** (sd. 38)	**193** (sd. 29)	**-2.60** (sd. 12)
	Male	**298** (sd. 48)	**305** (sd. 35)	**33** (sd. 15)	**86** (sd. 31)	**188** (sd. 22)	**-6.16** (sd. 12)
Babylon - Searching	Female	**323** (sd. 80)	**341** (sd. 58)	**45** (sd. 23)	**87** (sd. 38)	**175** (sd. 24)	**-0.29** (sd. 11)
	Male	**311** (sd. 81)	**343** (sd. 78)	**40** (sd. 21)	**82** (sd. 40)	**172** (sd. 24)	**-0.23** (sd. 11)
BBC - Searching	Female	**282** (sd. 44)	**338** (sd. 49)	**38** (sd. 18)	**82** (sd. 26)	**176** (sd. 21)	**-1.80** (sd. 10)
	Male	**270** (sd. 47)	**326** (sd. 53)	**42** (sd. 28)	**86** (sd. 44)	**176** (sd.23)	**-2.08** (sd. 14)
GoDaddy - Searching	Female	**392** (sd. 93)	**352** (sd. 59)	**68** (sd. 18)	**72** (sd. 22)	**208** (sd. 31)	**1.79** (sd. 13)
	Male	**378** (sd. 88)	**337** (sd. 57)	**64** (sd. 21)	**69** (sd. 21)	**215** (sd. 36)	**5.02** (sd. 15)
Yahoo - Searching	Female	**297** (sd. 51)	**430** (sd. 72)	**31** (sd. 25)	**115** (sd. 52)	**160** (sd. 26)	**-2.94** (sd. 12)
	Male	**282** (sd. 86)	**433** (sd. 85)	**105** (sd. 51)	**159** (sd. 21)	**126** (sd. 32)	**-1.44** (sd. 12)

Appendix O

DEMOGRAPHICS OF PARTICIPANTS IN THE VALIDATION STUDY

Table O.1: Demographics of Participants in the Validation Study

ID	Gender	Age-Group	Educational Level	GoDaddy	Apple	AVG	Yahoo	Babylon	BBC
1	Female	18-24	High School	5	5	5	1	5	4
2	Female	25-34	Undergraduate	4	1	5	1	5	5
3	Male	35-54	Graduate	5	4	5	1	5	3
4	Female	35-54	Middle School	5	5	5	5	5	5
5	Female	25-34	High School	4	4	4	1	3	3
6	Male	25-34	Graduate	5	3	5	4	5	2
7	Male	25-34	Undergraduate	5	3	3	1	4	2
8	Male	25-34	Undergraduate	5	1	3	1	1	1
9	Male	35-54	Undergraduate	4	5	5	5	5	5
10	Male	25-34	Undergraduate	5	3	4	1	5	1
11	Male	35-54	Undergraduate	5	2	4	3	5	3
12	Female	35-54	Graduate	5	1	2	3	4	5
13	Male	35-54	High School	5	4	3	4	5	3
14	Female	35-54	Undergraduate	5	4	5	3	5	4
15	Female	25-34	High School	5	4	5	4	5	4
16	Male	25-34	High School	2	1	3	1	5	5
17	Male	35-54	High School	5	3	5	4	5	5
18	Male	25-34	Undergraduate	4	4	5	4	4	1
19	Female	18-24	High School	5	2	5	2	1	2
20	Male	25-34	High School	4	1	4	4	5	1