



T. C.

Altinbaş University

Graduate School of Science and Engineering

Information Technology

**EMOTIONAL STATE CLASSIFICATION BY USING EVENT
RELATED POTENTIALS**

Shams Firas AL-ADHAMI

M. Sc. Thesis

Supervised by Prof. Dr. Oğuz BAYAT

Istanbul, 2019

**EMOTIONAL STATE CLASSIFICATION BY USING EVENT RELATED
POTENTIALS**

By

Shams Firas AL-ADHAMI

Information Technology

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science

ALTINBAŞ UNIVERSITY

2019

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Adil Deniz DURU

Co- Supervisor

Prof. Dr. Oguz BAYAT

Supervisor

Examining Committee Members

Prof. Dr. Oguz BAYAT	School of Engineering and Natural Sciences, Altinbas University	_____
Prof. Dr. Osman Nuri UCAN	School of Engineering and Natural Sciences, Altinbas University	_____
Asst. Prof. Dr. Muhammad ILYAS	School of Engineering and Natural Sciences, Altinbas University	_____
Prof. Dr. Hasan Hüseyin BALIK	Air Force Academy , National Defense University	_____
Asst. Prof. Dr. Adil Deniz DURU	Faculty of Sport Science, Marmara University	_____

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Oguz ATA

Head of Department

Prof. Dr. Oguz BAYAT

Director

Approval Date of Graduate School of
Science and Engineering: ____/____/____

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.

SHAMS FIRAS AL-ADHEMI

ACKNOWLEDGEMENTS

All praise and thanks are due to Allah S.W.T, and peace and blessing be upon his messenger Mohammed (peace be upon him).

Firstly, I would like to express my deepest thankfulness, admiration, and gratitude to prof.Dr. Adil Deniz DURU for his dedication, support, patience and guidance, with his knowledge he helped this study to see the light. Thank you so much.

Special thanks prof.Dr. Oguz BAYAT for his help through my studying time.

I would like to thank my parents, my Father **Firas Mehjin Aladhmi** and my Mother **Noor Laith Alrasheed**. I would not be able to achieve anything in life without them. Thank you for all the days that you spent it helping me through life. No words can express my gratitude and thankfulness to them, and to my little brother **Mohammed** the reason of my happiness and to all my family .

I would like to dedicate this effort and fatigue and this achievement to the spirit of my grandmother **Maeda** who did not leave me how I wished to be present to be proud of my achievement.

A very special word of thanks goes to my friends for their support and help through everything, in addition to ALmuatasim for his extraordinary support and help. I am so lucky to have a friends like you.

Finally, I would like to thank all staff of the faculty of Information Technology for their kindness and cooperation during my study.

SHAMS FIRAS AL-ADHEMI

ABSTRACT

Emotional State Classification By Using Event Related Potentials

AL-ADHAMI, Shams Firas ,

M.Sc. Information Technology, Altınbaş University,

Supervisor: Prof. Dr. Oğuz BAYAT

Co-Supervisor: Asst. Prof. Dr. Adil Deniz DURU

Date: 10/09/2019

Pages: 51

Human brain is a mystery that's been a curious subject to scientists for a long time. Therefore many studies took place regarding the subject of unfolding the reaction of the human brain to various stimulus using EEG and various techniques. In this study the reaction of the brain to the written words will be discussed using multiple classification techniques such as (KNN, MLP, SVM and naive bayes) classification algorithms. Where the emotional meaning words were collected through a survey and divided into positive meaning words and negative meaning words. After collecting the words the experiment took place where 10 males were subjected to the words and the collected data from the brain readings passed through the used classification techniques. The results showed that it is possible see the brain reaction to the words, in addition a comparison took place among the classification techniques to determine which one is the best for this type of problems. The SVM classification algorithm presented the best accuracy for training 98% and for testing 92%.

Keywords:

TABLE OF CONTENTS

	<u>Pages</u>
ABSTRACT	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiii
1. OVERVIEW	1
1.1 INTRODUCTION	1
1.2 LITERATURE REVIEW.....	3
1.3 OBJECTIVE OF THE STUDY.....	12
1.4 ORGANIZATION OF THE THESIS.....	12
2. EEG, ERP AND DATA MINING CLASSIFICATION TECHNIQUES.....	13
2.1 EEG.....	13
2.1.1 EEG Mechanism	13
2.1.2 EEG Advantages	14
2.1.3 EEG Disadvantages.....	15
2.2 ERPs (EVENT RELATED POTENTIALS).....	15
2.3 CLASSIFICATION METHODS	16
2.3.1 K-NN (K nearest neighbor)	17
2.3.2 Multilayer Perceptron (MLP)	19
2.3.3 Support Vector Machines (SVMs).....	21
2.3.4 Naïve Bayes	23
2.4 CROSS VALIDATION	26
2.5 CONFUSION MATRIX	26
2.6 DATA PREPROCESSING	28
2.6.1 Data Normalization	28
2.6.2 Min-Max Normalization.....	29
3. SYSTEM IMPLEMENTATION	30
3.1 PREPERATION	30
3.2 APPLYING KNN CLASSIFICATION ALGORITHM.....	37

3.3 APPLYING MULTILAYER PERCEPTRON (MLP)	38
3.4 APPLYING SUPPORT VECTOR MACHINE (SVM) CLASSIFICATION ALGORITHM.....	39
3.5 APPLYING NAÏVE BAYES CLASSIFICATION ALGORITHM	40
4.RESULTS	41
4.1 K-NEAREST NEIGHBOR (KNN) CLASSIFICATION RESULTS	41
4.2 MULTILAYER PERCEPTRON (MLP) CLASSIFICATION RESULTS	41
4.3 SUPPORT VECTOR MACHINE (SVM) CLASSIFICATION RESULTS	42
4.4 NAÏVE BAYES CLASSIFICATION RESULTS	42
4.5 ACCURACY COMPARISON.....	43
5.CONCLUSION AND FUTURE WORK	44
5.1 CONCLUSION	44
5.2 FUTURE WORK	44
REFERENCES.....	45

LIST OF TABLES

	<u>Pages</u>
Table 2.1: Confusion Matrix	27
Table 3.1: The Survey Words	30
Table 3.2: The Final Words	34
Table 4.1: Accuracy of KNN	41
Table 4.2: Results of KNN	41
Table 4.3: Accuracy of MLP	41
Table 4.4 Results of MLP	41
Table 4.5: Accuracy of SVM	42
Table 4.6: Results of SVM.	42
Table 4.7: Accuracy of Naïve Bayes	42
Table 4.8: Results of Naïve Bayes	43
Table 4.9: Accuracy Comparison between the Used Methods	43

LIST OF FIGURES

	<u>pages</u>
Figure 2.1: The Multi-Layer Perceptron Model	19
Figure 2.2: SVM Hyper Plane	22
Figure 3.1: Positive Emotion Word	35
Figure 3.2: Negative Emotion Word.....	35
Figure 3.3: Electrodes Placement in the EEG System	36
Figure 3.4: The EEG Signals	36
Figure 3.5: Example From The Used Dataset	37
Figure 3.6: K Value.....	38

LIST OF ABBREVIATIONS

EEG	Electroencephalography
ERP	Event-related potentials
KNN	K-nearest neighbor
MLP	Multilayer Perceptron
SVM	Support vector machines
FFT	Fast Fourier Transformation
ASI	Asymmetry Index
FMRI	Functional Magnetic Resonance Imaging
AB	Attentional improvements
EPN	Early posterior negativity
LPC	Later Centro-parietal
CW	Connected words
NCW	Non-connected words
RT	Reaction time
EMG	Electromyography
MEG	Magnetoencephalography

PET Positron Emission Tomography

SPECT Single-Photon Emission Computed Tomography

MRI Magnetic resonance imaging



1. OVERVIEW

1.1 INTRODUCTION

For the past decade the mystery of the human brain has been the main focus in so many studies, Research on this area is growing rapidly. By understanding the way that's the human brain works too many paths will be paved toward better real life developments. This study tries to understand the reaction of the brain toward feelings, specifically written words of feelings, furthermore the brain usually processes the written words in a regions like visual cortex, corticolimbic circuit and frontal regions, and the experiment has been done on native Arabic speakers specifically males. The aim of the study understand the human brain response to written word in addition to discuss the best classification algorithm that suitable for this type of case studies. In order to achieve the goal of the presented study which is showing how the brain of humans react to written words that contains an emotional meaning, EEG (electroencephalogram)/ERP (event related potentials) measurements are used.

EEG can be defined as the electrical action related to the alternating type that is recorded from the surface of the scalp following being picked up via conductive media and metal electrodes[1] [2], while ERP offers voltage measurement that is related to the neural activity which could be non-invasively recorded from multiple regions on the scalp [3], ERP have the ability of reflecting the summed postsynaptic potentials which have been produced in neural transmission and implemented passively via the skull and the brain to the surface of the skin in which they have the ability of contributing to EEG [4]. The data that been used in this study is collected from ten test subject (men), where the test subjects are exposed to a written word on a computer screen to evaluate different reaction of brain by using EEG/ERP. After collecting the data many classification techniques are used, classification can be defined as data analysis type which operate by extracting the models that are defining significant classes of data [5]. The used classification techniques are:

1. KNN (k nearest neighbor): can be considered as a major supervised learning algorithm that is used in pattern classification. The whole training set is retained via this rule throughout the learning process, also this process signify a class for each query, such class is specified via majority label regarding its KNN in training set [6].

2. MLP (multilayer perceptron): can be defined as a class that is related to the feed forward ANN. Minimum of 3 layers of nodes are included in MLP: input layer, hidden layer as well as output layer. Apart from input node, all nodes are considered as neurons which applies non-linear activation function. For the purpose of training, back propagation (that is a supervised learning method) is utilize by MLP.[7][8] MLP is distinguished from linear perceptron via nonlinear activation as well as its multiple layers. Data that is non linearly separable can be distinguished via MLP.[9]
3. SVM (support vector machine): is majorly derived from the binary classification problem. These 2 aspects are used to describe the major concept of SVMs: First, it works on constructing non-linear kernel function for presenting the inner product that is related to the feature space, which correspond to mapping data from input space to feature space of high-dimension via non-linear algorithm. Therefore, the samples' non-linear characteristics could be analyzed in feature space with a linear algorithm. Secondly, it applies a structural risk minimization principle in a statistical learning theory [10] through generalizing the optimal hyper-plane with the margin between the 2 classes is maximum. Even though that it is naturally straightforward, such concept is of high importance in capacity controlling, also it does make learned machine not just has minor risks, yet it also has optimum generalization performance. Thus, SVMs have various benefits in practical view and theoretical base [11].
4. Naïve Bayes: can be defined as a common method for classifying documents and can be powerful when using terms frequency. This method proved to be efficient when applied in a past studies, the consistent results of this classifier were also reported in studies concerning classification tasks[12].The naive Bayes classifier greatly simplify learning by assuming that features are independent given class. Although independence is generally a poor assumption, in practice naive Bayes often competes well with more sophisticated classifier. [13].

There are too many other classification algorithms that can be used for the same aim of this study but this study chose the above mentioned algorithms because it's expected from previous studies to give the best expected results.

In addition to the previous aim of the study a comparison between the proposed algorithms is going to be presented for the purpose of determining the best algorithm for this type of problem.

1.2 LITERATURE REVIEW

The human brain and its functionality has been a great mystery for scientists, for many years many studies tried to get a better understanding on this mystery here are some related studies to this topic

M. Chan et al. [14] applied machine-learning (ML) algorithms for the purpose of demonstrating that the individual words and the semantic category, could be decoded from the recording of MEG and EEG of a subject that is doing a language task. The mean accuracies of 76 percent (chance=50 percent) and 83 percent (chance=20 percent) have been acquired for decoding related to living vs. non-living category or individual words respectively. Also, the authors applied such decoding analysis for demonstrating that semantic category and the words' representations are extremely distributed temporally and spatially.

D.Rugg and E.Nagy [15] here the ERPs have been recorded as the subjects were doing two recognition memory tasks. With regard to the first task, sequence of words has been provided where the items have been repeated following six or nineteen intervening words. Old/new discrimination regarding each word has been required from the subjects, while in the second task, that has been carried out in about 45 minutes following the first task, the subjects were required to distinguish between the words which were used in first task and formerly un-presented items. With regard to first task, ERP to the 'old' words have been extra positive-going, from approximately 250 msec post stimulus, than those to the 'new' items. The old/new differences of ERPs in second task have been smaller than the first task, not as much as common on scalp, while the onset latency has been 500 msec, even though that the ability of the subjects for making 'old/new' discriminations has been at high-level. It is proposed that the 'old/new' differences of ERPs could not be reflecting the processes essential for the discrimination between new and old items in the recognition memory tasks.

A.Grant [16] the aim of this paper is examining the way a variety of accents are observed, and the way that this could be related to the electrical activity of the brain. In the study, 14 female members from Roanoke College listened to 12 recording of various accents in English language whereas EEG have been utilized for measuring their brain activity. There has been 2 recording for each of the 6 speakers. English was the native language to 4 of the 6 speakers while it has been the non-native language for 2 of them. Each one of participants listened to female and male

Saudi Arabia, Irish and American speaker. Following listening to all the recordings, participants have been requested to evaluate the accent on Likert scale for likability, comfort-level, understanding, and reliability, and likability. Electrodes have been set over frontal lobe for adequate understanding regarding the accents' executive reactions, in addition to over parietal lobe for measuring the basic audition. After that, Fast Fourier Transformation (FFT) have been utilized for analyzing the data for finding the brain waves' frequency. Considerable impacts related to the way the participants observed the various accents according to the survey data have been identified. The following observations have been identified; the Saudi Arabian accent was the least understood accent while the American was the most understood accent, the participants liked listening mostly to the Irish accent and they have higher rating for females. The results of the study indicated no considerable observations for the recording of EEG. Such results have indicated that additional studies should be carried out, restructuring while the participants are being asked questions that are related to accent and making a task throughout the recording for allowing the participants to have extra concentration on the text's actual content.

M.Teplan [17] presented a study that is related to EEG measurements. The main aim of the study is helping with the orientation in EEG as well as to build essential understanding related to carrying out the recording of EEG. There are 2 parts in this study. The first part presented certain research subjects related to EEG, historical overview and subject's background, while the EEG recordings will be explained in the second part. The research indicated that for applying EEG methods, these components of the recording system are needed:

1. Electrode cap that has a conductive jelly or the Ag-AgCl disc electrodes with the conductive paste.
2. Amplifiers that have common-mode rejection rate of 100 dB or more, input impedances of 100 M Ohms or more, and an overall amplification gain between 100-100,000
3. Analog filters which are integrated in unit with a high-pass filter with cut-off frequency in the range 0.1-0.7 Hz and low-pass filter with a cut-off frequency not more than one half of sampling rate. The frequencies over 50 Hz are infrequently used since they participate insignificantly to EEG's power spectrum.
4. No less than 12 bit A/D converter with precision under the overall noise (0.3-2 μ V pp.), and the sampling frequency typically between 128 – 1024 Hz.

5. Adequately fast computer to handle the data for recordings and finally for online analysis, with acceptable hard disc volume.
6. Digital high-pass FIR filter with comparable cut-off frequency as analog high-pass.

The overall quality related to the equipment of recording is decided by the suitable combination of stated parameters. Raw signal of EEG must be checked for artefacts prior to further data processing.

A. Turnip et al. [18] discusses a classification technique for emotion, inspired and based on back propagation neural networks, which is utilized for classifying EEG-SSVEP signals. EEG-SSVEP signals were produced by four stimulation short movies. The study found out that for a group of emotions varies from sad to calm and (angry, happy) in between, the obtained accuracy was 75%. In addition it was proposed that in order to increase the accuracy of EEG, a more precise emotional design should be provided. The study presents the idea that an EEG-SSVEP signals can be used for communication of emotion of user as BIC. In addition the study shows that according to the obtained results, the experiment paradigm can be improved in addition to signal processing algorithm enhancement.

Xiao-Wei Wang et al. [19] proposes a comparison among three types of EEG's existing features for the classification of emotions. In addition it proposes a smoothing technique (LDS) for noise cancelation to remove unwanted emotion task, lastly it proposes a simple method for emotion trajectory changes detection and tracking. In addition the study shows that the classification accuracy can be improved significantly by using LDS smoothing technique. This conclusion conducted from the results of the emotion classification by using SVM classifier. Among the three types of feature represented in the study, the most robust one is the power spectrum for all the bands of frequencies. In addition the results shows that the lower frequencies are less important than the higher frequencies regarding the emotion activities.

C. Petrantonakis and J. Hadjileontiadis [20] presented a new approach to evaluate emotion elicitation processes in EEG-based emotion recognition setup. Through the use of frontal brain asymmetry theory, an index (asymmetry Index (AsI)), has been presented for the purpose of evaluating such asymmetry. This is achieved via multi-dimensional directed information analysis between that various EEG sites from 2 opposite brain hemispheres. The suggested method has been utilized for 3-channel (Fp1, Fp2, and F3/F4 10/20 sites) recording of EEG taken from sixteen healthy right-handed participants. For evaluating the effectiveness of AsI, general

classification procedure has been implemented through the use of 2 feature-vector extraction approaches and SVM classifier for 6 distinctive classification situations in valence/arousal space. This led to a classification results equal to 62.58% for user independent situation, 94.40% for user-dependent situation, proving the effectiveness related to AsI as index for the emotion elicitation assessment.

Stemmer and F. Connolly [21] offered an outline regarding major software development and hardware system of EEG/ERP presently on the market in addition to the advantages of this technology for studying issues related to language issues. The research examine the “added-value” technology and what it bring to language research and the potentials to combine different neuro-imaging approaches with the importance on integrating EEG/ERP and the functional magnetic resonance imaging (fMRI). The paper concludes with indicating the methodologies that improve the area in the future.

Sharavani Sur and V.K.Sinha [22] indicates how an optimum medium is provided by EEG for understanding the neuro-biological deregulation with possibility of evaluating neurotransmission, time-locked electroencephalogram activity or the ERPs help in capturing the neural activity an outline related to various waveforms regarding ERPs and the main findings in different psychiatric situations. Also, the paper indicated that ERPs establishes millisecond-by-millisecond processing of neural information that could be related to specific operation like updating the working memory and encoding inhibitory responses. Therefore, it offers non-invasive methods for evaluating the brain function in individuals experiencing cognitive disorders and it is considered to be of prognostic value in certain conditions, ERP can be considered as an approach of neuropsychiatric research that is anticipated highly in the future.

J. Luck et al. [23] studies a few of the current ERPs, concentrating on researches that separate operation of attention in certain cognitive sub-systems like response selection, working memory and perception. A few conclusions have been observed. The first conclusion is that within certain situations attention modulate initial feed forward volley regarding the neural activity in intermediate visual processing. The second conclusion is that such early impacts could be identified for automatic capture of attention and the voluntary allocation of attention after peripheral visual transient. The third conclusion is that such impacts exists not just in the case when the attention is directed to location in 2-D space, however also in the case when the attention is directed to either of two spatially overlapping surfaces. The fourth conclusion is that

the sensory activity is not modulated via the attention if not the sensory systems have been overloaded; in the case when the sensory systems are not taxed, attention could function to impact the memory or response processes. The attention functions for mitigating the information overload in either cognitive sub-systems overloaded via specific combination of task and stimuli. T. Koeniga et al. [24] presented a research that indicate the various neural activities throughout the mental imagery, abstract mentation could be given to clearly identified brain information processing steps. Throughout a randomized visual presentation that is related to single, abstract-type and imagery-type words, twenty-seven channel ERP field maps have been acquired from twenty-five subjects (the sequence-divided to first and second group for statistics). Brain field map series indicated sequence related to general map configuration which have been quasi-stable for the brief time periods (microstates). Microstates have been concatenated via rapid changes in the map. Since various map configurations should be derived from various spatial patterns of neural activities, different active neural network is represented by each microstate. So, the microstates are expected to agree with certain discrete steps related to information-processing. Associating the microstate topographies (through the use of centroids) between the abstract-type and imagery-type words, considerably distinctive microstates have been detected in the two subject groups at 286–354 ms in which the imagery-type words have been more right-lateralized than the abstract-type words, and at 550–606 ms and 606–666 ms in which the anterior-posterior differences happened. The research found that the processing include some clearly identified stages and that the brain-states including such stages are changed via stimuli's capacities for generating mental imagery or abstract mentation in state-dependent method.

E. M. Sokhadze et al. [25] highlighted and explored the measuring event-related potential (ERP) usefulness in cognitive test. In addition to discussing the ERP application. The paper explains how the ERP is still as important as before for the brain research in the field of psychiatry. The (ERP) data based markers is very beneficial to researchers because it will give an in depth view to the psychophysiology of psychopathologies. In addition the ERP method is considered relatively cheap so it is accessible to many individuals.

M. Adamo and S. Ferber [26] indicated the way that 2 streams are interacting by consolidation and attentional selection in a way that the objects whose form fits function regarding formerly perceived object, like a tool, take advantage of attentional improvements because of the tool's action affordance. Through the use of ERP and attentional blink (AB) paradigm, the researchers

examined if offering action relation between 2 objects looking closely together in time counteracts the general decrement detected for processing of second item. They utilized names (experiment 1b) and images (experiments 1a and 2) of general tools, objects which could be acted upon via such tools, and irrelevant objects. They identified the fact that the pictorial presentation regarding a tool as well as its action matching part lead to decreased attentional blink in addition to improved attentional selection perceived as larger P3, relative to irrelevant objects and tools, also such attentional improvement does not depend on semantic associations. This implies that action affordance instantiated through tool's perception is going to decrease the functional blindness generally identified in the case when 2 targets are very close in time, particularly when the tool's physical features which elicit an action affordance are observed.

Marta Kutas and Kara D. Federmeier [27] indicated the detection, description, and developing usage of N400, an ERP brain response related to the meaning processing. They defined the elicitation regarding N400 via notable variety of stimulus types such as spoken, written, and pseudo or signed words; photos, videos of faces, objects, drawings, and actions; sounds; and mathematical symbol as well as outlining the sensitivity related to N400 amplitude (since its latency is constant) to non-linguistic and linguistic manipulations. They highlighted the efficiency of N400 as important variable for studying nearly all aspects related to language processing and draw attention to its increasing applications to the probe semantic memory and for determining the way neurocognitive systems easily and dynamically applies top-down and bottom-up information to make sense of the world. They underlined the various theories related to the functional importance of N400 and provide an N400-inspired reconceptualization regarding the way that meaning processing could be clarified. Furthermore, ERPs offer temporally superb brain measures which proved to be efficient tools to get the features related to cognition mainly impenetrable with the behavioral measures to distinguish among qualitatively distinctive procedures and track their time courses, regularly with no overt task and occasionally in the nonexistence of awareness. ERP data in most cases and the N400 especially highlight the important role of time (and timing) in understanding (as in cognition more commonly). N400 was involved in breaking the barriers between the domains. N400 literature, offer convincing picture related to the way language, attention, perception, and memory cooperatively contribute to the neural actions responsible for N400 and therefore be involved in such astonishing capability of human brain to infuse its world with meaning.

Harald T. Schupp et al. [28] indicated that the selective visual attention is guided via the emotional pictures. In which, series of studies related to ERPs are showing the robust and consistent modulation regarding certain components of ERP through emotional images. Particularly, pictures that are showing natural un-pleasant and pleasant scenes related to elevated early posterior negativity, late positive potential, as well as sustained positive slow wave in comparison to neutral contents. Such modulations are believed to index various phases of stimulus processing such as stimulus representation in the working memory, perceptual encoding, , and elaborate stimulus evaluation. Also, the research involve a discussion related to papers examining the collaboration regarding the motivated attention with active and passive forms of attentional control. Resent study is considered to be examining selective processing regarding the emotional cues as function of stimulus novelty, in context of explicit attention tasks, learned stimulus significance and emotional prime pictures. It is indicated that the measures of ERPs are of high importance in evaluating emotion–attention interface at level of different processing phases. The results have been discussed as part of 2-phase models of stimulus perception obtained from examining learning, orienting and attention.

Johanna Kissler et al. [29] The main aim of this paper has been emotional words are specially processed throughout silent reading through the use of ERP. An evaluation will be conducted regarding we investigate to what degree various components that are related to visual evoked potential, that is P1, N1, early posterior negativity (EPN, about 250 ms following word onset) in addition to late positive complex (LPC, about 500 ms) react differentially to the emotional words and if such response is based on the accessibility of certain attentional resources. Subjects observed random sequences related to unpleasant, neutral and pleasant nouns and adjectives. Initially, they have been trained to only read words and after that count nouns or adjectives. There has been no consistent impacts occurred for P1 and N1. Yet, throughout counting and reading, EPN has been improved for the emotionally arousing words (unpleasant and pleasant), in spite of if the word have its place in non-target or target category. Task effect on EPNs has been limited to the adjectives, yet the impact has no interaction with the emotional content. Later centro-parietal LPC (450–650 ms) indicated a great improvement for attended word class. Topographically distinct and small emotion- late positive complex effect has been identified particularly responding to the pleasant words, throughout active tasks and silent reading. Therefore, the emotional word content has been processed with ease and in automatic way, also

it is not subjected to interfere from primary grammatical decision task. The results of the study came in accordance with other studies related to early automatic semantic processing as replicated via posterior negativities in ERPs about 250 ms following word onset. Suggestions regarding the models of the brain's emotion–attention interactions have been examined.

H. Taha et al. [30] examined the impacts of orthographic connectivity on time course that is related to early brain electric response throughout visual word recognition. To achieve this task, ERPs has been collected from adult practiced readers as carrying out lexical decision task through the use of fully connected (Cw), partially connected and non-connected words (NCw). The reactions time variance has been higher and the precision has been lower in NCws in comparison to Cws. ERP analysis indicated considerable amplitude and latency variations between NCws and Cws at the posterior electrodes throughout N170 component that indicated temporo-occipital areas. The results of the study indicated that rather than slowing the reading, the orthographic connectivity in practiced readers of Arabic impacts the reading positively throughout the early phases of word recognition. Such results have been examined with regard to former notes in literature.

Detlef E. Dietrich et al. [31] this research has the aim of delineating the impact of emotional content that is related to the stimuli to be considered on the recognition performance related to normal subjects through ERP. In the case of repeatedly presenting the words, the brain will response repeated and identified items are considered by a waveform of extra positivity ('old/new effect'). Words have been considered for the emotional connotation they have ('neutral', 'negative' and 'positive') have been specified continually on video monitor to the subjects, that have the task of indicating if a certain word happened for first ('new') or the second ('old') time in the list through pressing one of the two buttons. With regard to each of the word categories, ERP that have old words have been more positive in comparison to those of new items from approximately 250 ms following stimulus. Old/ new impact has been considerably improved for positive and negative items in comparison to neutral stimuli between 450 and 650 ms following stimulus indicating a considerable impact related to word's emotional content on the processes of verbal memory. The results of the study proved that the item's emotional content has a vital impact on the recognition's electrophysiological concomitants. The model seems to be possible for examining differential impact related to the psychotropic

compounds on the cognitive and emotional procedures and for researches related to the altered cognition in the mood disorders.

Susan J. Thomas et al. [32] indicated the existence of preferential processing that is related to the threatening information according to reaction time (RT) impairment in the competing task. Due to the fact that the measures of ERP are considered to have sensitivity to speed (latency) and extent (amplitude) of the cerebral processing, they have been useful tools that are used to study the idea that threatening stimuli are related to the enhanced attention. Words' pool have been rated by 22 students for identifying the ones that were personally distressing. There are 2 types of words (neutral and threat) have been put to comparison in 2 tasks (color related, where words' color ink has been recognized, and word relevant where the words have been organized as threatening or not threatening). Emotional Stroop effect has not been identified with regard to longer RTs for identifying colors related to the threat words. The results of ERPs offered helpful information regarding the threat processing that has not been identified with the behavioral measures. The threat content has been related to larger P2 amplitude in right than the left hemisphere, as well as larger P3 amplitude, through tasks. The results of the study indicated solid proof for improved processing associated to the threat-related stimuli in the healthy subjects. The conclusion is that ERP is sensitive measure for processing the underlying emotional Stroop performance, that could be applied for elucidating the attentional biases with regard to clinical and healthy populations.

Joshua D. Davis et al. [33] indicated theoretically and dynamic significant discussion related to when and how the embodiment contribute to the language comprehension. The discussion tackled such issues via examining the interference with the facial actions effects the real time response of the brain to the emotional language. The participants read sentences related to certain negative and positive events (e.g., "She reached inside the pocket of her coat from last winter and found some (cash/bugs) inside it.") whereas ERP has been recorded. The facial actions of the participants has been influenced through requesting the participants to take chopsticks in their mouth through the use of a position which blocked or allowed smiling, as established via EMG. To block smiling was not so effective in ERP to valenced words (bugs, cash) however it did impact ERP to the sentence's final words defining positive events. The results of the study have indicated that adequately positive sentences have the ability of evoking smiles and this facial action could be facilitating semantic processing indexed via N400 component. Generally, the

study provides common indication that that embodiment effects certain features related to the high-level comprehension, including constructing situation model.

1.3 OBJECTIVE OF THE STUDY

The main aim of the presented thesis is summarized in the following way:

1. Showing the brain response to the emotional written words.
2. Implementing KNN classification algorithm.
3. Applying MLP classification algorithm.
4. Calculate the results of applying SVM classification algorithm.
5. Implementing naïve bayes classification algorithm.
6. Comparing the results of the used algorithm and determining which one is the best for this problem.

1.4 ORGANIZATION OF THE THESIS

Chapter 1: In this chapter, a short introduction related to EEG/ERP and classification algorithms will be presented, in addition this chapter includes literature review, objective of the study and organization of the presented study.

Chapter 2: In this chapter, the details of EEG/ERP will be discussed in detail, in addition to the classification algorithms and all the components that been used in this study.

Chapter 3: In this chapter, the dataset and the implemented system will be described.

Chapter 4: In this chapter, the results of the study will be presented.

Chapter 5: In this chapter, the conclusions and future work will be presented.

2. EEG, ERP AND DATA MINING CLASSIFICATION TECHNIQUES

2.1 EEG

Encephalography has experienced great developments throughout over one-hundred years of its history. In the year 1875, the English physician Richard Caton discovered the presence of brain's electrical currents, he identified EEGs from exposed brains of monkey and rabbit. In the year 1924, they German neurologist Hans Berger utilized his radio equipment for amplifying the electrical activity of the brain that was measured on the scalp of human. He indicated that the weak electric currents that are created in brain could be recorded with no need to open the skull and it could be shown graphically on a paper. Depending on the brain's functional status, there was a change in the activities he detected, including lack of oxygen, anaesthesia and sleep and in particular neural diseases, like epilepsy. Berger paved the way for various current applications related to electroencephalography, also he used the term electroencephalogram for defining the electric potentials of human brain. He was accurate with his idea that the activity of the brain changes in detectable and consistent manner in the condition when the subject's general status changes, such as from relaxation to awareness [2].

The scalp recordings that are related to the brain's spontaneous neuronal activity, recognized as the electroencephalogram, allow measuring possible modifications in time between reference electrode and signal electrode [34]. In comparison to the other biopotentials like ECG, the EEG is very complex for untrained observer to interpret, partly due to the spatial mapping related to the functions on various regions of brain and electrode location. Identifying that certain standardization has been required, International Federation in Electroencephalography and Clinical Neurophysiology implemented 10-20 electrode placement system, and the study applied sixteen electrode [35].

2.1.1 EEG Mechanism

With regard to the traditional scalp EEGs, the recordings have been acquired through using a conductive paste or gel for placing electrodes on scalp, typically after getting the scalp area ready for reducing impedance because of the cells of dead skin. Electrodes are used by a lot of systems, each one of the electrodes is attached to a single wire. Nets or caps are used by few systems into which the electrodes are embedded; such systems are general in the case when high-density arrays of electrodes are required. The international 10-20 system is used to specify the names and

locations of electrodes [36]. Each one of the electrodes will be connected to single input regarding differential amplifier (single amplifier for each two electrodes); general system reference electrode is connected to other input of each differential amplifier. Such amplifiers function by amplifying voltage between reference and active electrode (generally 1,000–100,000 times, or 60–100 dB of voltage gain). With regard to the analog EEGs, the signal will be filtered (next paragraph), EEG signals will be the output as the deflection related to the pens while the paper pass beneath. Recently, the majority of EEGs are digital, the amplified signals will be digitized through analog-to-digital converters and data will be displayed on monitors, following passing through anti-aliasing filter. The analog-to-digital sampling generally applied at 256–512 Hz in the clinical scalp EEGs; the sampling rate of up to 20 kHz are applied in certain study applications. Throughout the recordings, certain activation processes will be applied. Such processes will make abnormal or normal EEG activities which could not else be perceived. Such processes involve sleep deprivation and sleep, eye closure, hyperventilation, mental activities, and photic stimulation (with strobe light). Throughout the (inpatient) epilepsy monitoring, typical seizure medications related to the patient could be taken. Digital EEG signals are being stored electronically and could be filtered for displaying. General properties for high-pass filters and a low-pass filters are 0.5–1 Hz and 35–70 Hz respectively. Such high-pass filters generally filter out the slow artifacts, including movement artifact and electrogalvanic signals, while low-pass filters operate by filtering high-frequency artifact, like electromyographic signal. Additional notch filters are usually applied for removing resulted from electrical power lines (60 Hz in United States and 50 Hz in other nations) [37].

2.1.2 EEG Advantages

The costs for hardware are considerably less than those related to the majority of other approaches [38]. EEGs prevent the limited availability related to technologists for providing instant care in high-traffic hospitals [39]. The sensors of EEGs could be applied in more places than MEG, PET, fMRI, MRS, or SPECT, since such approaches need immobile and bulky. For example, MEG involves equipment that consist of liquid helium-cooled detectors which could be utilized in magnetically shielded rooms, overall costing few million dollars, and fMRI needs using 1-ton magnet in, again, a shielded room [40]. EEGs has extremely high temporal resolutions, on order of milliseconds instead of seconds. EEGs are generally recorded at

sampling rate between 250 and 2000 Hz in the research and clinical properties, however, recent EEG data collection systems have the ability of recording at sampling rate over 20,000 Hz if needed. EROS and MEG are the only other non-invasive cognitive neuroscience approaches that obtain data at such level of temporal resolution [40]. EEGs are fairly tolerant regarding the subject movement, dissimilar to the majority of other neuroimaging approaches. Some approaches exist to minimize and eliminate the movement artifacts in EEG data [41]. EEG does not aggravate claustrophobia, unlike fMRI, PET, MRS, SPECT, and sometimes MEG [42]. EEGs do not include exposure to the high-intensity (>1 tesla) magnetic fields, like in certain other approaches, particularly MRS and MRI. These have the ability of resulting in various unwanted problems with data, also prohibit using such approaches with participants having metal implants in their body, like metal-containing pacemakers [43]. EEGs do include the exposure to radioligands, dissimilar to the positron emission tomography [44]. EEG have the ability of detecting the covert processing (processing which does not need response) [45]. EEG could be applied in subjects that do not have the ability to make motor response [46]. Few components of ERP could be identified even in the case when the subject does not attend the stimuli Unlike the other methods to evaluate RT, ERPs could elucidate the stages related to processing (instead of only final end result) [47]. EEG can be defined as an efficient tool to track brain changes throughout various stages of life. The sleep analysis of EEG could specify considerable features related to the timing of the brain development, such as estimating the adolescent brain maturation [48].

2.1.3 EEG Disadvantages

The low spatial resolution on scalp. For example, fMRI could show brain areas which are active, whereas EEG needs powerful interpretation for hypothesizing the areas which are via specific response [49]. Unlike PET and MRS, do not have the ability of identifying certain brain location at which different drugs and neurotransmitters could be detected [44]. Poor signal-to-noise ratio, thus the sophisticated data analysis and fairly large numbers of subjects are required for extracting important information from EEG [50].

2.2 ERPs (Event Related Potentials)

The ability of human beings in feeling, thinking, and acting might somehow be a result of the brain operations. For more than 100 years, scientists utilized brain activity measures for gaining

information about the cognitive, motor, and perceptual functions. Therefore, scholars developed many different approaches for non-invasively measuring brain activity. Those approaches are classified to two categories: “electromagnetic” methods which measure the activity of the brain directly with recording electromagnetic fields that are produced by particular populations of neurons, and “hemodynamic” methods which measure the activity of the brain indirectly with recording variations in the vascular variables which are associated with the neural activity changes. Significantly, those approaches vary in several aspects, which include pre-conditions for signal detection, the homogeneity with which the activity of neurons is sampled from various brain parts, and the relative strengths in the determination when versus where neural activity happens. Which is why, they provide complementary views on the neural activity [51]. In an earlier study which involves those brain potential measures, the term ‘evoked potential’, or EP, has been deployed due to the fact that it was assumed that possibilities represented the activity of the brain which has been strictly ‘evoked’ with the presentation of the stimulus, activity which is associated with fundamental sensory processes [52]. ERPs are quite small voltages that are produced in the structures of the brain as a response to certain stimuli or event. Those are EEG changes which are time locked to sensory, cognitive or motor events providing non-invasive and safe mechanism of studying psycho-physiological mental processes correlates [22].

Event Related Potentials in humans may be classified to 2 categories. The early ways or components which peak roughly within the first 100 milliseconds after stimuli, are referred to as ‘sensory’ or ‘exogenous’ as they highly depend on the stimulus of the physical parameters. As opposed, ERPs that are produced in later parts reflect how the subject evaluates the stimulus and are referred to as cognitive or ‘endogenous’ Event Related Potentials as they view the processing of information [22].

2.3 CLASSIFICATION METHODS

Classification is a data analysis type, which obtains models that describe valuable classes of data. Those models, which are referred to as classifiers, are capable of predicting categorical class labels (i.e. discrete, unordered) [5]. In another words classification can be defined as the methods that is used for the classification of every item in a dataset to a predetermined set of groups or classes [53]. Data classification is a process that includes two steps, which are a learning step in which a model of classification is produced and a step of classification, in which the model is

utilized for predicting class labels for certain data [5]. The aim of classification is the accurate prediction of the aimed class for every case in the data [53].

The approaches that are most commonly utilized for tasks of data mining classification are categorized to the following classes: [53]

1. Rule-based approaches.
2. Methods of decision tree induction.
3. Neural networks.
4. Memory based learning.
5. SVMs.
6. Bayes network.

This study used the following classification techniques:

2.3.1 K-NN (K nearest neighbor)

K-NN algorithm is a very simple classification algorithm. In spite of its simplicity, it is capable of giving rather competitive results. K-NN algorithm may as well be utilized for problems of regression [54]. K-NN algorithm can memorize all of the training data and can perform classifications only in the case where the test attributes of the object are exactly matching to one of the training samples. An evident disadvantage of this method is that numerous test records are not going to be classified due to the fact that they don't match any training record exactly. A more advanced method is the k-NN classification [55], this approach obtains a set of k objects in the training set which are nearest to the test object, and assigns a label based on how predominant a certain class in this neighborhood. There are three main factors in this method: a group of labeled objects, such as a measure of distance or similarity for the calculation of the distances between the objects, a set of stored records, the k value, as well as the number of nearest neighbors. For the sake of classifying unlabeled objects, it is required to calculate the distance from a given object to the labeled objects, then, identifying its k-nearest neighbors, finally, the class labels of those nearest neighbors are utilized for determining that object's class label [56].

Considering and a test object $x = (x', y')$ and a training set D , the algorithm performs a calculation of similarity (or distance) between z and each training object $(x, y) \in D$ for the sake

of determining the list of its nearest-neighbors, D_z . (x Is the training object data, whereas y is the class. In a similar manner, x' is the testing object data and y' is its class). Right after obtaining the list of the nearest-neighbors, a classification of the test object is performed on the basis of the nearest neighbors' majority class:

$$\text{Majority Voting: } y' = \arg \max_v \sum_{(x_i, y_i) \in D_z} I(v = y_i) \quad (2.1)$$

Here, v represents a class label, y_i represents the i th nearest neighbors' class label, and $I(\cdot)$ is an indicator function returning the value 1 in the case where its argument is true and 0 if its argument is not true [56].

K-NN classification technique can be presented with the following steps:

Input: D , the group of k training objects, and test object $z = (x', y')$

Begin:

Calculate $d(x', x)$, which is the distance between z and each one of the objects, $(x, y) \in D$

Choose $D_z \subseteq D$, the group of k nearest training object to z .

Output: $y' = \arg \max_v \sum_{(x_i, y_i) \in D_z} I(v = y_i)$

As shown in previous studies K-NN is especially good for multi-modal classes and also in applications where one object might have multiple class labels. For instance, for assigning functions to genes according to the profiles of expression, some of the scholars have discovered that K-NN has performed better than the SVM, which is a considerably more advanced approach of classification [57].

2.3.1.1 Advantages of KNN

Its robustness to noisy training data and its effectiveness in the case of large training datasets, Simplicity, Effectiveness and Competitive performance of classification in a wide range of domains Intuitiveness [58].

2.3.1.2 Disadvantages of KNN

It has high sensitivity to redundant or irrelevant features due to the fact that each feature contributes to the similarity, and consequently to the classification. This might be overcome by

careful feature weighting or selection. K-NN might have insufficient run-time performance in case of a large training set. Costs of computation are rather high due to the need for computing the distance of every one of the queries rather than to all of the training samples. Distance based learning isn't clear what distance type to utilize and what feature to utilize for producing the optimal results [58].

2.3.1.3 Predictions of k-nearest neighbor

After the selection of k value, predictions can be made based on K-NN examples. For regression, K-NN predictions is the mean value of the result of the k-nearest neighbors.

$$y = \frac{1}{K} \sum_{i=1}^K y_i \quad (2.2)$$

Where y_i denotes the *ith* case of the examples sample and y denotes the prediction (i.e. the result) of query point. As opposed to the regression, in tasks of classification, predictions of K-NN are based on a voting approach where the winner is utilized for labeling the query [59].

2.3.2 Multilayer Perceptron (MLP)

A multilayer preceptor is a of feed forward ANN class. It is made up of a minimum of 3 node layers, which are: input, hidden and output layers. Every node, except the input ones, is a neuron which utilizes a non-linear function of activation. MLP uses a learning approach which is referred to as the back propagation (BP) for training. Its non-linear activation and several layers are distinguishing it from linear preceptors. It is capable of distinguishing data which isn't linearly separable [60].

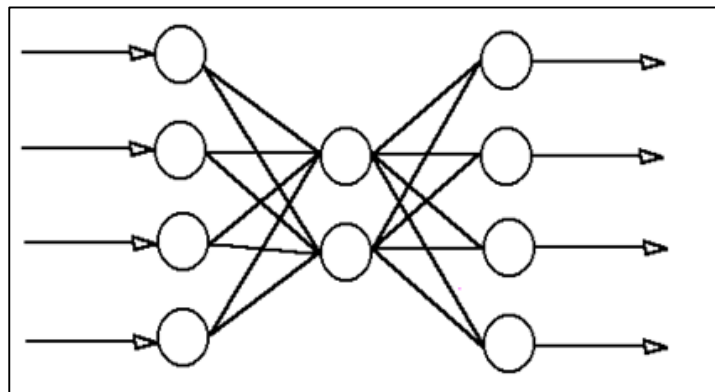


Figure 2.1 : The Multi-Layer Perceptron Model

For the sake of being capable of solving non-linearly separable tasks, multiple neurons are linked in layers for building a multi-layer perceptron. Every perceptron is utilized for the identification of small input sections that are linearly separable. Perceptron outputs are combined to another perceptron for producing the ultimate outcome. The hard-limiting (step) function which is utilized for the production of the output prevents data on the real inputs to flow on to the inner neurons. For solving this issue, the hard-limiting function is replaced by a continuous function, which is typically the *sigmoid function* [60]. Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result. This is an example of supervised learning, and is carried out through backpropagation, a generalization of the least mean squares algorithm in the linear perceptron.

We can represent the degree of error in an output node j in the j th data point (training example) by

$$e_j(n) = d_j(n) - y_i(n) \quad (2.3)$$

Where d is the target value and y is the value produced by the perceptron. The node weights can then be adjusted based on corrections that minimize the error in the entire output, given by

$$\mathcal{E}(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (2.4)$$

Using gradient descent, the change in each weight is

$$\Delta w_{ji} = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n) \quad (2.5)$$

where y_i is the output of the previous neuron and η is the learning rate, which is selected to ensure that the weights quickly converge to a response, without oscillations. The derivative to be calculated depends on the induced local field v_j , which itself varies. It is easy to prove that for an output node this derivative can be simplified to

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n) \dot{\phi}(v_j(n)) \quad (2.6)$$

Where $\dot{\phi}$ is the derivative of the activation function described above, which itself does not vary. The analysis is more difficult for the change in weights to a hidden node, but it can be shown that the relevant derivative is

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = \dot{\phi}(v_j(n)) \sum_K -\frac{\partial \mathcal{E}(n)}{\partial v_k(n)} w_{kj}(n) \quad (2.7)$$

This depends on the change in weights of the k th nodes, which represent the output layer. So to change the hidden layer weights, the output layer weights change according to the derivative of the activation function, and so this algorithm represents a backpropagation of the activation function [61].

2.3.2.1 Advantages of MLP

Fault Tolerance, This property is referred to as “graceful degradation” as well. Due to the fact that it has distributed nature, a NN continue operating even in the case of the failure of a significant portion of its interconnections and neurons. In addition to that, re-learning post damage may be rather fast [60]. **Generalisation**, NN are able generalize, which means that they are capable of classifying an unknown pattern with other known patterns sharing identical distinguishing characteristics. Which indicates incomplete or noisy inputs will undergo classification due to having similarity with complete and pure inputs [60].

1.3.2.2 Disadvantages of MLP

It has the issue of scaling also It is considered a computationally expensive learning process and no guaranteed solution [60].

2.3.3 Support Vector Machines (SVMs)

The Classification algorithm which is based on SVM now drives more attention because of its perfect theoretical characteristics and sufficient empirical outputs [11]. An SVM is a supervised algorithm of machine learning which may be utilized for the purposes of regression and classification. SVM is more often utilized in tasks of classification [62]. It is based on the

concept of detecting a hyper plane which best splits a data-set to two classes, as depicted in the following figure [62].

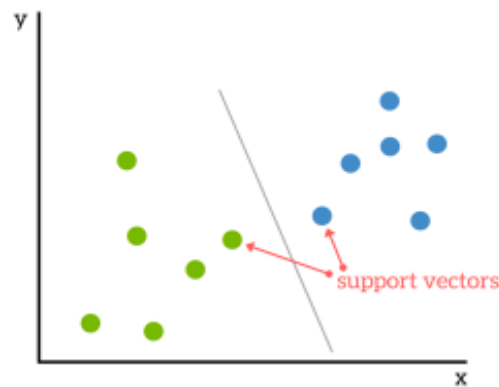


Figure 2.2 : SVM Hyper Plane

SVMs are data points which are closest to the hyper plane, the dataset points that, if eliminated, would update the dividing hyper plane position. As a result, they may be considered as the critical dataset elements [62].

The goal of SVMs is finding the optimal function of classification for distinguishing between members of both classes in training data. The measure for the idea of the “optimal” function of classification may be accomplished geometrically. For a linearly separable data-set, a linear function of classification is corresponding to a separating hyper-plane $f(x)$ which passes by the middle of both classes, as a result separating them. As soon as this function is found, new data sample x_n may be categorized via simply testing the sign of the function $f(x_n)$; x_n is part of the positive class if $f(x_n) > 0$. Due to the fact that there is a wide range of such linear hyper-planes, the thing SVM additionally guarantees is that the optimal function is determined by the maximization of the edge between both classes. This edge is characterized as the amount of separation or space between both classes as characterized by the hyper-plane. In a geometrical sense, the edge indicates the smallest distance between data points which are nearest to a point on the hyper-plane. In light of this geometrically expressed definition, we are allowed to explore the way of maximizing the margin, in a way that although there are infinitely many hyper-planes, only a limited number of them qualifies as a solution for the SVM. The reason behind SVMs insisting on determining the maximal margin hyper-planes is the fact that it presents the optimal ability of generalization. It permits not only the optimal performance of classification (accuracy, for example) on the training data, but as well permits precise classification of future data. For the

sake of ensuring that the maximal margin hyper-planes are, in fact determined, a classifier of SVM makes the attempt of maximizing the following function according to b and \vec{w} [56]:

$$L_p = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^t \alpha_i y_i (\vec{w} \cdot \vec{x}_i + b) + \sum_{i=1}^t \alpha_i \quad (2.8)$$

Where t represents the number of training samples, $\alpha_i, i = 1, \dots, t$, are non-negative numbers in a way that L_p derivatives according to α_i are 0. α_i represent the Lagrange multipliers and L_p is referred to as the Lagrangian. In this formula, the constant b and the vectors \vec{w} define the hyper-plane [56].

2.3.3.1 Advantages of SVM

SVMs are excellent when there is no information about the data. It operates efficiently even with even semi structured and unstructured data such as images, trees, and text. In contrast to the NNs, SVMs are not solved for local optima. The kernel trick is the actual strength of the SVMs. With a suitable kernel function, there is a possibility of solving any complicated task. Models of SVM have generalization in practice, the risk to over-fit is smaller in the SVMs. It scales rather sufficiently to data of high dimensions [63].

2.3.3.2 Disadvantages of SVM

It is not easy to elect a “good” kernel function. There is a difficulty in understanding and interpreting the ultimate model, individual impact, and variable weights. It takes a long time to train large data-sets. The hyper parameters of SVMs are gamma and Cost -C. It isn’t that easy fine-tuning those hyper-parameters, and difficult to envision their effect. Due to the fact that the final model isn’t so easy to see, there is no ability for doing small adjustments to the model, which is why, it is difficult incorporating specific business logic [63].

2.3.4 Naïve Bayes

The Naïve Bayesian classification method is greatly appealing due to its elegance, robustness, and simplicity. It is one of the earliest algorithms of formal classification, however, even in the simplest form it has, it is usually greatly efficient [56]. The most powerful solutions are often the simplest ones, and Naïve Bayesian proves that. Despite the major Machine Learning advances that happened in the past years, it has proved to be simple as well as time efficient, reliable, and accurate. It is utilized for a wide range of application, but works quite efficiently with problems

of natural language processing (NLP). Naïve Bayesian is a group of probabilistic algorithms which benefit from the Bayes' Theorem and probability theory for the prediction of the text tag (such as a customer review or piece of news). They're probabilistic, meaning that they compute the probability of every tag for a certain text, and after that, they give the tag that has the maximum probability. The way they obtain those probabilities is with the use of Bayesian Theorem, describing a feature's probability, according to prior knowledge about conditions which could be associated with that feature [64].

Naïve Bayesian classification can be defined as an approach which applies to a specific class of tasks, which are the ones which may be referred to as relating an object to a discrete category [65].

The Naïve Bayesian theorem works by letting D be a training set that includes tuples with the class labels that are related to them. As usual, every one of the tuples is denoted with an attribute vector of n -dimensions, $X = (x_1, x_2, \dots, x_n)$, which depicts n measurements that are made on the tuple from n attributes, respectively, A_1, A_2, \dots, A_n . And assuming that there are m classes, C_1, C_2, \dots, C_m . Considering a tuple, X , the classifier is going to predict that X is part of the class that has the highest posterior possibility, conditioned on X . which means that, the Naive Bayes classifier predicts that tuple X is part of the class C_i only in the case where [5]

$$P(C_i|X) > P(C_j|X) \text{ for } 1 \leq j \leq m, j \neq i. \quad (2.9)$$

Which is why, $P(C_i|X)$ the class C_i for which $P(C_i|X)$ is maximized, is referred to as the *maximum posteriori* hypothesis. By Bayesian theorem which is

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \quad (2.10)$$

Where $P(X)$ is constant for every class, $P(X|C_i) P(C_i)$ only requires maximizing. In the case where the prior probabilities of the class are unknown, then it is usually presumed that classes have equal likelihood, which means, $P(C_1) = P(C_2) = \dots = P(C_m)$, and thus, we would maximize $P(X|C_i)$. Or else, $P(X|C_i) P(C_i)$ would be maximized. It should be noted that the prior probabilities of a class could be predicted by $P(C_i) = |C_i, D|/|D|$, where $|C_i, D|$ denotes the number of training tuples of class C_i in D . for the sake of reducing calculation in the evaluation of $P(X|C_i)$, the naïve assumption of class-conditional independence is done. Which assumes that the values of attributes are conditionally independent of each other, considering the tuple's class

label (in other words, the fact that there isn't any dependence correlations amongst attributes).
Therefore:

$$P(X|C_i) = \prod_{K=1}^n P(X_K|C_i) \tag{2.11}$$

$$P(X|C_i) \times P(X_2|C_i) \times \dots \times P(X_n|C_i)$$

There is a possibility of easily estimating probabilities $P(X_1|C_i)$, $P(X_2|C_i)$, ..., $P(X_n|C_i)$ from training tuples. Recalling that X_K in the equation above indicates attribute A_K value for the tuple X . For every one of the attributes, we examine if the attribute is continuous-valued or categorical. For example, for computing $P(X|C_i)$, the following aspects are considered [5]:

1. The case where A_K is categorical, then $P(X_K|C_i)$ is the number of class C_i tuples in D which have the value X_K for A_K , divided by $|C_i, D|$, the number of class C_i in D tuples.
2. The case where A_K is continuous-valued, in this case there is a need for doing a little bit more work, however, the computations are rather straight-forward. Continuous-valued attributes are typically expected to have a Gauss distribution with a standard deviation σ and an average value μ that are defined as [5]:

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.12}$$

So that

$$P(X_K|C_i) = g(X_K, \mu_{C_i}, \sigma_{C_i}) \tag{2.13}$$

2.3.4.1 Advantages of naïve bayes classifier

It has high level of effectiveness (i.e. filters over 99.691%). It is conceptually quite easy to comprehend. Numerous Email clients nowadays are either directly or indirectly supporting Bayes filtering. Everyone's filter is basically customized, which makes it quite hard for spammers defeating each individual's filter with a one message [65].

2.3.4.2 Disadvantages of naïve bayes classifier

The process of initialization is rather time consuming (however, this issue may be overcome via caching word counts for every folder of the mail). There is a need for having an amount of bad and good messages for initializing the filter. On every one of the messages, a user-specific database of word likelihoods must be consulted. Which makes the Bayes filtering rather resource intensive and maybe unideal for sites that have large user databases. There is an occurrence (however, rare) of false positives [66].

2.4 CROSS VALIDATION

In k-fold cross-validation, in some cases referred to as estimation of rotation, the data-set D is split in a random manner to k sub-sets (i.e. folds) which are mutually exclusive D_1, D_2, \dots, D_K and have almost equal sizes. The inducer will be trained and tested k number of times, every time $t \in \{1, 2, \dots, k\}$, it will be trained on D/D_t and tested on D_t . The cross-validation accuracy estimation is the entire number of precise classifications, divided by the number of samples in the data-set. In other words, $D_{(i)}$ is the testing set including instance $x_i = \langle v_i + y_i \rangle$, which means that the cross-validation estimate the precision:

$$acc_{cv} = \frac{1}{n} \sum_{(v_i, y_i \in D)} \delta(I(D/D_{(i)}, v_i), y_i) \quad (2.14)$$

The estimation of the cross-validation is an arbitrary number which is dependent on dividing to folds. The entire cross-validation is the mean value of all $\binom{m}{m/k}$ likelihoods for selecting m/k samples from m , however, it's often very costly. Except for the leave-one-one (n-fold cross-validation) that is typically complete, k-fold cross-validation estimates full k-fold cross-validation with the use of one split of data to folds [67].

2.5 CONFUSION MATRIX

It includes data concerning expected and actual classifications that are made with a system of classification. The efficiency of this type of systems is often assessed with the use of data which is in the matrix. The table below exhibits confusion matrix for a 2-class classifier [68].

The variables in the confusion matrix mean the following:

1. a represents the number of correctly made predictions of a sample being negative.
2. b represents the number of wrong projections of a sample being positive.
3. c represents the number of incorrectly made predictions of a sample being negative.
4. d represents the number of correctly made predictions of a sample being positive.

Table 2.1: Confusion Matrix

		Predicted	
		negative	Positive
actual	negative	a (TN)	b (FP)
	positive	c (FN)	d (TP)

A wide range of standard terms were characterized for two-class matrices:

1. accuracy (AC): which indicates the rate of the entire number of correctly made predictions. Its value is calculated with the use of equation (2.15):

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.15)$$

2. The recall is the rate of positive instances which have been identified correctly, the value if recall is obtained with the use of equation (2.16):

$$recall = \frac{TP}{TP + FN} \quad (2.16)$$

3. precision (P): this value indicates the rate of predicted positive instances which have been correct, precision can be computed with the use of equation (2.17):

$$precision = \frac{TP}{TP + FP} \quad (2.17)$$

4. The support: it is the number of true response samples which reside in that class.

5. The f1-score: it provides the harmonic average value of the recall and precision. The scores which correspond to each one of the classes will indicate the classifier's precision in the classification of data points in that certain class, in comparison with the rest of the classes [68].

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (2.18)$$

2.6 DATA PREPROCESSING

Pre-processing is sometimes a neglected however, important step in the process of data mining. Data collection is typically a procedure which is lightly regulated, and that results in values that are out of range, for instance, illogical combinations of data, missing values, and so on. The analysis of data which hasn't been checked thoroughly for this kind of issues may result in incorrect outcomes. Which is why, the quality of data and its representation is firstly and most importantly to maintain prior to executing an analysis. In the case where there is a big amount of redundant and irrelevant data or noisy and unreliable data, this will result in the knowledge discovery to be more complicated to perform. Preparing data preparation could take a long time to perform. Pre-processing includes preparing the data, which is accompanied with data integration, normalization, cleaning, and transformation; in addition to tasks of data reduction; like feature selection, discretization, instance selection, and others. The result which is expected following a reliable chaining of data pre-processing phase is a final data-set that may be taken under consideration as accurate and useful for additional algorithms of data mining [69].

2.6.1 Data Normalization

Data normalizing the attempt of data in giving each attribute the same weight. This process is especially useful in algorithms of classification which involve NNs or distance measurements like clustering and nearest-neighbor classification. In the case of utilizing neural network back-propagation approach for classification mining, the normalization of the input values for every one of the attributes that is measured in training tuples is going to be helpful in speeding up the process of learning. For approaches that are distance-based, normalization can be helpful in the

prevention of attributes that have initially large ranges from out-weighting the attributes that have initially smaller ranges. In addition to that, it is useful in the case where there is no advance insight about the data [5].

There are numerous approaches of data normalization. Like the min-max, z-score, in addition to the normalization with decimal scaling. For the present discussion, assume that A is a numerical attribute that has n observed values, v_1, v_2, \dots, v_n . [5]

2.6.2 Min-Max Normalization

This type of normalization carries out a linear transform on original data. Assuming that min_A is the minimum value of an attribute and max_A is its maximum value, A. Min-max maps of normalization a value, v_i , of A to v'_i in range $[new_min_A, new_max_A]$ via calculating:

$$v'_i = \frac{v_i - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A \quad (2.19)$$

This type of normalization maintains the correlations amongst the original values of data. It will meet an “out-of-bounds” error in the case where the future case or input for the normalization falls outside the original range of data for A [5][69].

3. SYSTEM IMPLEMENTATION

The purpose of the study is to show the brain reaction to a written emotional words specifically in Arabic language and discussing how a different classification algorithms will perform and compare between the results to see which method will give the best results accuracy wise.

3.1 PREPERATION

The main idea of the study is to show an Arabic words (with different meanings) on a screen in front of 10 Arabic speaking the test subjects (males). The words were selected by making a survey about a number of words that presented in the table below:

Table 3.1: The Survey Words

NO	YES	The words in Arabic language	The words meaning in English language
3.2%	96.8%	النجاح	Success
25.8%	74.2%	النوم	Sleep
11.4%	88.6%	الرضا	Satisfaction
24.9%	75.1%	الحب	Love
11.6%	88.4%	الام	Mother
46.6%	53.4%	الرقص	dance
9.4%	90.6%	المطر	Rain
17.4%	82.6%	الصديق	Friend
13.3%	86.7%	السفر	Travel
22.5%	77.5%	الحجاب	Hijab
48.1%	51.9%	اللون الاسود	The color black
36.7%	63.3%	الوطن	Home
84.1%	15.9%	الاركيلة	Hookah
56.9%	43.1%	الاقارب	Relatives

13.5%	86.5%	المال	Money
87.3%	12.7%	التردد	Hesitation
16.5%	83.5%	الثقة	Trust
89.9%	10.1%	الاشمئزاز	Disgust
86.9%	13.1%	الغرور	Ego
57.9%	42.1%	الواقع	Reality
11.4%	88.6%	الهدوء	Quite
24.7%	75.3%	شوكولاتة	Chocolate
63.3%	36.7%	المجاملة	Complement
34.8%	65.2%	القدر	Destiny
92.7%	7.3%	الاكتئاب	Depression
84.5%	15.5%	الانتقام	Revenge
25.5%	74.5%	الاحلام	Dreams
6.7%	93.3%	الجمال	Beauty
86.3%	13.7%	الخوف	Fear
7.5%	92.5%	التواضع	Humbleness
57.9%	42.1%	النسيان	Forgetting
67.0%	33.0%	الصيف	Summer
99.1%	0.9%	النفاق	Hypocrisy
15.2%	84.8%	الاطفال	Kids
25.5%	74.5%	المستقبل	Future
55.4%	44.6%	اللون الاصفر	The color yellow
12.2%	87.8%	الطعام	Food
21.9%	78.1%	المزاح	Joking
92.5%	7.5%	الطلاق	Divorce

97.0%	3.0%	العنصرية	Racism
15.7%	84.3%	الموسيقى	Music
31.8%	68.2%	مواقع التواصل الاجتماعي	Social media
24.9%	75.1%	الدين	Religion
76.2%	23.8%	الفضول	Curiosity
98.5%	1.5%	الحسد	Envy
88.8%	11.2%	الغضب	Anger
13.3%	86.7%	الطيبة	Kindness
8.4%	91.6%	الامل	Hope
8.7%	91.3%	المسؤولية	Responsibility
15.3%	84.7%	البحر	Sea
3.0%	97.0%	الاخوة	Brotherhood
14.8%	85.2%	الرياضة	Sport
75.6%	24.4%	الالم	Pain
4.6%	95.4%	النظام	Order
58.2%	41.8%	الانتقاد	Criticism
75.6%	24.4%	التوتر	Stress
93.6%	6.4%	الكذب	Lie
1.4%	98.6%	الامانة	Honesty
92.5%	7.5%	الاستغلال	Exploitation
44.1%	55.9%	العتاب	Repentance
88.1%	11.9%	الفشل	Failure
1.8%	98.2%	السلام	Peace
32.2%	67.8%	الخجل	Shyness
94.3%	5.7%	الاهانة	Insult

89.7%	10.3%	الكره	Hatred
62.8%	37.2%	السياسة	Policy
61.6%	38.4%	الغموض	mystery
5.5%	94.5%	الاصرار	Determination
18.3%	81.7%	الذكريات	Memories
94.1%	5.9%	الرشوة	Bribe
86.1%	13.9%	القسوة	Harshness
17.8%	82.2%	الاشتياق	Missing
8.4%	91.6%	التضحية	Sacrifice
17.6%	82.4%	اللون الاخضر	The color green
41.3%	58.7%	التقدم بالعمر	Aging
8.4%	91.6%	الالتزام	Commitment
7.8%	92.2%	العاطفة	Passion
32.2%	67.8%	النصيب	The share
88.4%	11.6%	التصنع	Fakeness
2.5%	97.5%	الادب	Politeness
58.9%	41.1%	الندم	Regret
5.5%	94.5%	السعادة	Happiness
23.1%	76.9%	الزواج	Marriage
5.9%	94.1%	الصبر	Patience
5.9%	94.1%	المرح	Joy
40.6%	59.4%	الماضي	The past
53.9%	46.1%	الدموع	Tears
4.3%	95.7%	الوفاء	Loyalty
5.0%	95.0%	اللقاء	Meeting

1.6%	98.4%	الكرامة	Dignity
63.0%	37.0%	الجنون	Craziness
85.4%	14.6%	الذنوب	Sins
5.7%	94.3%	الايمان	Faith
11.6%	88.4%	الضيف	Guest
81.3%	18.7%	التباهي	Show off
6.4%	93.6%	الدراسة	Study
50.0%	50.0%	اللون الاحمر	The color red
72.1%	27.9%	الغربة	Alienation
75.3%	24.7%	الغيرة	jealousy

The above mentioned words were used in a survey where the survey asked 1000 people which of these words are the most effective emotionally and the results showed that 10 positive meaning words and 10 negative meaning words were the most effective and used in this study. The final words is shown in the following table:

Table 3.2: The Final Words

Positive meaning words		Negative meaning words	
Words in Arabic	Words in English	Words in Arabic	Words in English
النجاح	Success	التردد	Hesitation
الرضا	Satisfaction	الالم	pain
الحب	Love	الكذب	Lie
الثقة	Trust	الكره	Hatred
الامل	Hope	الاستغلال	Exploitation
الخجل	Shyness	الاهانة	Insult
الاشتياق	Missing	القسوة	Harshness
الذكريات	Memories	الغضب	Anger

المرح	Joy	الغيرة	Jealousy
الطيبة	kindness	النفاق	hypocrisy



Figure 3.1: Positive Emotion Word (love)



Figure 3.2: Negative Emotion Word (pain)

The test subjects were 10 native Arabic speaking males with age range of (25-40) years old. These test subjects were exposed to the final words shown above on a computer monitor while each test subject were connected to an EEG system that contain 16 electrodes (F3,Fp1,P3,O1,C3,Fz,T7,CZ,Fp2,F4,C4,T8,Pz,P4,O2,Oz) in a quite environment with no movement while the experiment were performed, the quite place and no movement were necessary to prevent distractions from the test subjects and achieve accurate results.



Figure 3.3 : Electrodes Placement In The EEG System

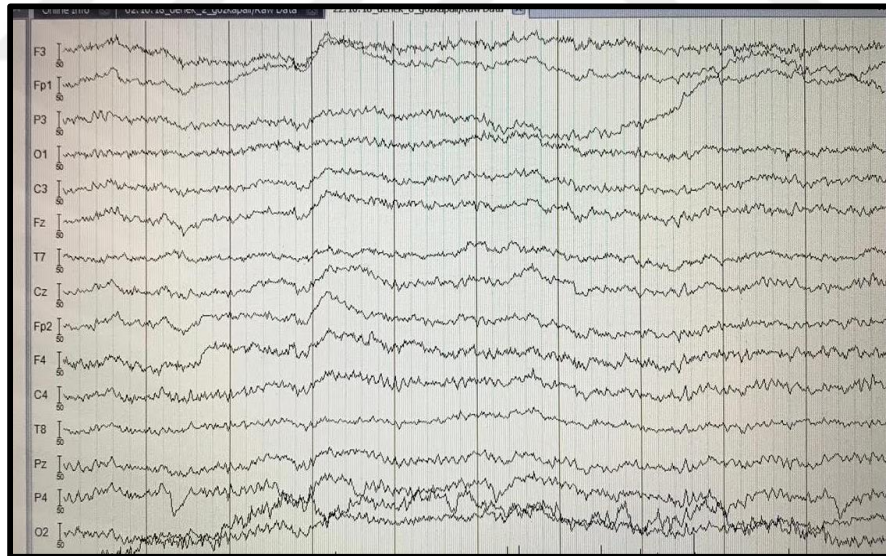


Figure 3.4: The EEG Signals

After collecting the EEG signals the ERP took place, the ERP data obtained from the time period Of (250-350) ms for the negative and positive peaks of the waveform.

After collecting the dataset of the two determined classes (positive emotions and negative emotions) that shown in the figure below, the preprocessing process start to normalize the data, for this study min-max method is used as a normalization method using the equation (2.19), this

method will help to achieve more organized data which in turn will increase the overall accuracy and improve the results.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	3.317369	21.05213	17.12437	18.48581	8.579791	15.58629	16.42171	-4.97264	10.85049	17.61841	4.579837	12.05058	12.39052	17.15914	10.23736	3					
2	-5.6154	2.731314	12.87324	-0.18468	7.292642	-0.39247	11.69201	-6.2338	10.404	17.99167	8.689235	6.776034	16.31494	23.42739	27.83657	3					
3	10.06685	16.80121	15.18774	12.49087	10.92059	3.810213	12.85821	7.696508	7.257198	8.834349	7.367564	24.82263	21.48507	24.78092	29.78336	3					
4	-5.72543	0.8702	5.871987	-0.90066	5.819556	1.062428	19.46344	-2.31404	3.207432	0.580016	1.496995	5.048052	8.025877	18.40276	7.575888	3					
5	7.841126	-0.49278	8.127012	-1.93107	3.359127	4.99336	26.57122	11.27074	7.307124	7.258583	3.234366	3.547668	3.743661	-3.09749	11.93847	3					
6	5.117127	7.400906	15.36618	-0.86846	10.73032	11.17039	14.52695	18.7321	16.98577	19.145	15.28807	10.94131	11.73114	15.688	21.18662	3					
7	3.432114	15.8569	13.97188	-2.15268	-1.45651	9.957977	-2.99484	4.627119	0.942188	12.10434	2.709188	11.25951	11.73676	6.20945	25.00015	3					
8	13.49032	9.060323	19.10998	4.473	13.01923	14.36238	5.642222	7.281881	2.773783	12.30327	5.268439	12.43507	1.287849	17.04345	24.31338	3					
9	0.786706	20.49593	23.74424	9.75006	4.49349	12.77933	14.00997	-0.93963	2.912151	20.09493	11.63245	23.49681	20.14384	32.8828	36.49473	3					
10	23.28917	5.50312	7.042361	2.26065	23.16652	7.967949	12.66344	24.67125	17.26963	4.936283	6.164891	0.858467	4.168885	13.99898	22.78722	3					
11	8.53588	5.400744	6.435306	10.13457	-7.1485	10.61958	3.393989	1.219885	4.221656	2.759267	-3.57073	1.733507	5.800474	11.70647	15.68314	3					
12	9.170903	24.01341	28.82286	19.94619	14.89158	8.226109	28.95651	10.07489	14.64492	38.44981	19.26578	31.88703	35.18217	31.38467	26.58959	3					
13	20.37341	7.665251	5.684797	20.76446	25.63415	4.925423	16.50347	14.87569	16.26967	18.82538	13.63003	9.955788	8.981227	0.567826	13.86664	3					
14	14.80124	19.93974	15.91726	21.78918	14.83104	10.91536	19.4947	13.17496	7.690847	16.45391	12.09575	27.11629	27.38202	25.15112	10.76221	3					
15	13.52176	18.13363	23.87195	22.46728	23.53695	7.05446	23.7964	21.25826	22.72733	39.2567	14.65024	27.11187	29.08104	29.60889	21.57374	3					
16	10.1926	14.09511	27.2956	14.35281	11.16418	12.2856	7.53218	15.24809	8.843439	23.94606	9.147774	10.83083	8.730214	21.50315	19.17655	3					
17	11.95778	8.849458	10.79665	12.87538	23.50236	11.12754	23.693	14.99279	22.25374	25.4436	11.32147	18.67746	25.76542	17.24012	17.21674	3					
18	10.85277	17.53771	18.66912	26.89956	29.42209	5.831596	24.37108	6.368167	13.43099	19.87142	28.43973	22.06051	28.85626	19.73122	30.80142	3					
19	6.997048	20.58303	13.59575	6.003461	9.127513	15.4964	24.0657	6.151256	11.93747	20.22233	7.848316	25.56508	24.45603	30.008	24.39155	3					
20	11.73929	3.507542	9.313131	-7.17972	15.25913	12.57198	17.32822	19.28878	22.14248	19.30593	19.91789	3.563136	8.020258	22.85089	14.49944	3					
21	21.56801	3.891072	-7.92233	11.53244	22.78743	6.021819	16.35198	11.84661	18.19542	17.22502	32.83949	8.344919	10.4461	-7.73843	5.234538	3					
22	14.30611	7.553707	3.247831	10.79373	6.087651	9.075843	10.86966	10.11324	13.93881	14.13161	38.45302	9.64864	18.24812	29.97137	10.69521	3					
23	19.39887	12.30734	7.432485	20.99554	10.55448	1.785694	22.51238	14.88529	11.06446	12.11328	9.804643	12.3732	8.053976	7.991015	1.186495	3					

Figure 3.5 : Example From The Used Dataset

In order to do apply the classification methods a 10 fold cross validation is applied to the normalized dataset. The 10 fold cross validation will divide the dataset into 10 groups of data, this way the dataset can be divided into training data and testing data . in this study the training data was 80% from the total data and the testing data was 20% from the total data.

The next step in the study is to apply the classification methods by using python programming language, the used version of the python computer program is 3.6.5.

3.2 APPLYING KNN CLASSIFICATION ALGORITHM

as shown in the previous chapter the KNN algorithm compares the data and assign the nearest data to the correct class, for this study k is determined as 7 because it has the highest as shown in figure 3.6 .

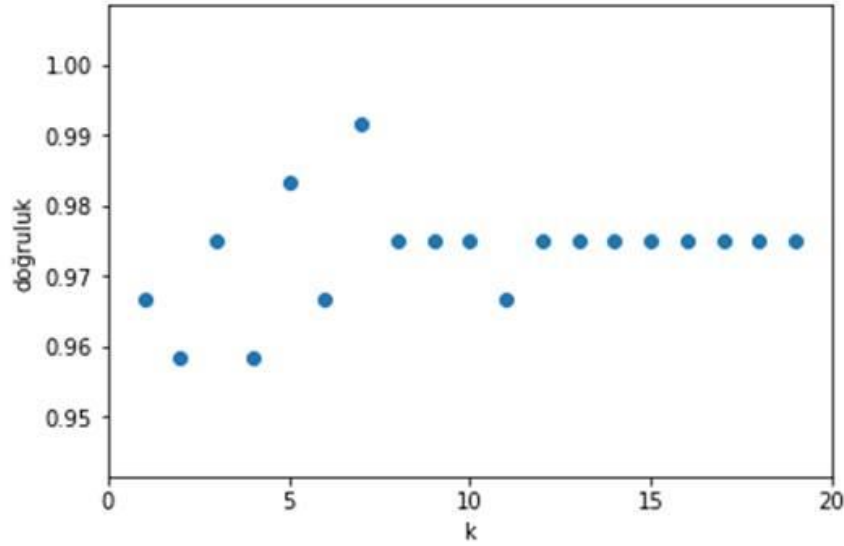


Figure 3.6 : K Value

value so it will compare 7 neighbors . This process is done by the Equation (2.1):

$$\text{Majority Voting: } y' = \arg \max_v \sum_{(x_i, y_i) \in D_z} I(v = y_i) \quad (2.1)$$

After applying the KNN algorithm, confusion matrix will take place to compute the accuracy, precision, F1 score and support using the following Equations:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.11)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (2.12)$$

3.3 APPLYING MULTILAYER PERCEPTRON (MLP)

As shown in the previous chapter the MLP classification algorithm, it changes the data into binary attributes where the data passes through the determined layers input layer, hidden layer and output layer. The MLP algorithm is applied using the equations

$$e_j(n) = d_j(n) - y_i(n) \quad (2.3)$$

$$\mathcal{E}(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (2.4)$$

$$\Delta w_{ji} = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n) \quad (2.5)$$

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n) \dot{\phi}(v_j(n)) \quad (2.6)$$

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = \dot{\phi}(v_j(n)) \sum_K -\frac{\partial \mathcal{E}(n)}{\partial v_k(n)} w_{kj}(n) \quad (2.7)$$

After applying MLP to the dataset and collecting classification results, the confusion matrix will be applied to produce the required measurements such as accuracy, precision, recall, F1 score and support.

Similar to all classification methods that used in this study it used 80% of the dataset as training data and 20% as a testing data.

3.4 APPLYING SUPPORT VECTOR MACHINE (SVM) CLASSIFICATION ALGORITHM

Support vector machine is as described in chapter two, classify the data by generating a hyper plane that separate the data the dataset into two classes as seen in Figure (2.2). In this study the SVM algorithm is used to classify the dataset into positive emotions and negative emotions and maximizing the margin between the hyper plane and the support vector. This can be achieved by implementing the following Equation:

$$L_p = \frac{1}{2} \|\bar{w}\| - \sum_{i=1}^t \alpha_i y_i (\bar{w} \cdot \bar{x}_i + b) + \sum_{i=1}^t \alpha_i \quad (2.4)$$

When the classification is done, the confusion matrix will be applied to get the accuracy, recall, precision, F1 score and support.

3.5 APPLYING NAÏVE BAYES CLASSIFICATION ALGORITHM

As the previous chapter implemented that the Naïve Bayes classification works on probability, it means that it works on a prior knowledge using the Equation:

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \quad (2.6)$$

So from the above equation the probability is computed and the dataset is classified into the required classes which is positive emotions and negative emotions.

After achieving the classification portion, the confusion matrix will produce the required elements for the test study results such as accuracy, recall, precision, F1 score and support.

4. RESULTS

This chapter will present the results of the used classification methods, the results will be introduced as accuracy, precision, recall, F1score and support.

4.1 K-NEAREST NEIGHBOR (KNN) CLASSIFICATION RESULTS

For the KNN classification algorithm the following tables presents the achieved results:

Table 4.1 : Accuracy Of KNN

Accuracy of training set	0.97
Accuracy of testing set	0.92

The above table shows the accuracy of the KNN classification algorithm for the training data which was 80% of the total data and for the testing data which was 20% from the total data.

Table 4.2: Results of KNN

	Precision	Recall	F1 score	support
training	1.00	0.17	0.29	6
testing	0.92	1.00	0.96	56

Table 4.2 presents the precision, recall, F1score and support for the KNN classification algorithm.

4.2 MULTILAYER PERCEPTRON (MLP) CLASSIFICATION RESULTS

For the MLP classification algorithm the following tables presents the achieved results:

Table 4.3 : Accuracy of MLP

Accuracy of training set	0.99
Accuracy of testing set	0.85

The above table shows the accuracy of the MLP classification algorithm for the training data which was 80% of the total data and for the testing data which was 20% from the total data.

Table 4.4: Results of MLP

	Precision	Recall	F1 score	support
training	1.00	0.17	0.25	6
testing	0.92	1.00	0.90	56

Table 4.4 presents the precision, recall, F1score and support for the KNN classification algorithm.

4.3 SUPPORT VECTOR MACHINE (SVM) CLASSIFICATION RESULTS

For the classification algorithm the following tables presents the achieved results:

Table 4.5: Accuracy of SVM

Accuracy of training set	0.98
Accuracy of testing set	0.92

The above table shows the accuracy of the Naïve Bayes classification algorithm for the training data which was 80% of the total data and for the testing data which was 20% from the total data.

Table 4.6: Results of SVM

	Precision	Recall	F1 score	support
Training	1.00	0.17	0.29	6
Testing	0.92	1.00	0.96	56

Table 4.6 presents the precision, recall, F1score and support for the KNN classification algorithm.

4.4 NAÏVE BAYES CLASSIFICATION RESULTS

For the (Naïve Bayes) classification algorithm the following tables presents the achieved results:

Table 4.7: Accuracy of Naïve Bayes

Accuracy of training set	0.97
Accuracy of testing set	0.90

The above table shows the accuracy of the (Naïve Bayes) classification algorithm for the training data which was 80% of the total data and for the testing data which was 20% from the total data.

Table 4.8: Results of Naïve Bayes

	Precision	Recall	F1 score	support
Training	0.50	0.17	0.25	6
Testing	0.92	0.98	0.95	56

Table 4.8 presents the precision, recall, F1score and support for Naïve Bayes the classification algorithm.

4.5 ACCURACY COMPARISON

The following table will present a comparison between the achieved accuracies from the classification algorithms that been used in this study, in addition the table will show the best method for the training dataset and for the testing set, furthermore it shows the best method among the used ones

Table 4.9: Accuracy Comparison between the Used Methods

Methods	Training Accuracy	Testing Accuracy	Best method
KNN	0.97	0.92	
MLP	0.99	0.85	
SVM	0.98	0.92	
Naïve Bayes	0.97	0.90	
Best result	0.99	0.92	SVM

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This study discusses different classification techniques (KNN, MLP, SVM and Naïve Bayes) classification algorithms to determine the brain emotional reaction to the written words, where the written words were in Arabic language.

The words were selected through a survey that done on a 1000 person, the final words were presented on a monitor to a 10 test subject which were native Arabic speakers (males).

By using EEG the brain signals are collected and converted to a numerical dataset.

The results of the classification showed that it is possible to classify the emotions with the used techniques, where the emotions were either positive emotions or negative emotions with different accuracies.

For the accuracies, the MLP classification algorithm showed the best training result with 99% accuracy while both KNN and Naïve Bayes presented the lowest training result with 97% accuracy. It should be noted that the training data is 80% of the total dataset.

For the testing portion of the dataset which were 20% of the total data the SVM and KNN classification algorithm showed the highest accuracy of 92%. Furthermore MLP classification algorithm showed the lowest accuracy with 85%

It can be concluded that it is possible to determine the brain emotion to a written words. Also it can be concluded that the SVM classification technique is the best technique to solve this type of classification problems, because it showed the highest testing accuracy with high training accuracy while other techniques showed lower accuracies.

5.2 FUTURE WORK

1. Expand the number of classification algorithms by including other techniques.
2. Adding other languages to see the brain reaction to different languages.
3. Increase the number of test subjects.
4. Include females in the test to compare the emotional reaction to the male brain.
5. Adding more emotions to the classification.
6. Using spoken words in addition to the written words.

REFERENCES

- [1] K. Keuper *et al.*, “How ‘love’ and ‘hate’ differ from ‘sleep’: Using combined electro/magnetoencephalographic data to reveal the sources of early cortical responses to emotional words,” *Hum. Brain Mapp.*, vol. 35, no. 3, pp. 875–888, 2014.
- [2] F. H. . O. T. Niedermeyer, Ernst . Lopes da Silva, *Electroencephalography: basic principles, clinical applications, and related fields* /, 5th ed. Philadelphia ; London : Lippincott Williams & Wilkins, c2005.
- [3] N. Birbaumer, T. Elbert, A. G. Canavan, and B. Rockstroh, “Slow potentials of the cerebral cortex and behavior.,” *Physiol. Rev.*, vol. 70, no. 1, pp. 1–41, Jan. 1990.
- [4] Lutzenberger, W., Elbert, T. H., & Rockstroh, B. *A brief tutorial on the implications of volume conduction for the interpretation of the EEG.* *Journal of Psychophysiology*, 1, 81-89, (1987).
- [5] M. Teimouri *et al.*, *Data Mining Concepts & Techniques*, vol. 15. 2016.
- [6] J. Gou, L. Du, Y. Zhang, and T. Xiong, “A new distance-weighted k-nearest neighbor classifier,” *J. Inf. Comput. Sci.*, vol. 9, no. February, pp. 1429–1436, 2012.
- [7] S. García, *Intelligent Systems Reference Library 72 Data Preprocessing in Data Mining*. 2015.
- [8] D. E. Ruineihart, G. E. Hint, and R. J. Williams, “LEARNING INTERNAL REPRESENTATIONS BERROR PROPAGATION two,” 1985.
- [9] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Math. Control. Signals, Syst.*, vol. 5, no. 4, pp. 455–455, Dec. 1992.
- [10] VAPNIK and V. N., “The Nature of Statistical Learning,” *Theory*, 1995.
- [11] R. Xiao, J. Wang, and F. Zhang, “An approach to incremental SVM learning algorithm,” *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI*, vol. 2000-January, no. 1, pp. 268–273, 2000.
- [12] S. Das and A. K. Kolya, “Sense GST: Text mining & sentiment analysis of GST tweets by

- Naive Bayes algorithm,” in *Proceedings - 2017 3rd IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2017*, 2017.
- [13] I. Rish, “An empirical study of the naive Bayes classifier,” pp. 41–46, 2010.
- [14] A. M. Chan, E. Halgren, K. Marinkovic, and S. S. Cash, “Decoding word and category-specific spatiotemporal representations from MEG and EEG,” *Neuroimage*, vol. 54, no. 4, pp. 3028–3039, 2011.
- [15] M. D. Rugg and M. C. Doyle, “Event-Related Potentials and Recognition Memory for Low- and High-Frequency Words,” *J. Cogn. Neurosci.*, vol. 4, no. 1, pp. 69–79, 2008.
- [16] A. Grant, R. Benons, A. Johns, M. Hobson, and D. Nichols, “Foreign Accent Perception and Processing with EEG,” pp. 1–9, 2016.
- [17] S. M. Teplan (Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, “Fundamentals of Eeg Measurement,” *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.
- [18] A. Turnip, A. I. Simbolon, M. Faizal Amri, P. Sihombing, R. H. Setiadi, and E. Mulyana, “Backpropagation neural networks training for EEG-SSVEP classification of emotion recognition,” *Internetworking Indones. J.*, vol. 9, no. 1, pp. 53–57, 2017.
- [19] X. W. Wang, D. Nie, and B. L. Lu, “Emotional state classification from EEG data using machine learning approach,” *Neurocomputing*, vol. 129, pp. 94–106, 2014.
- [20] P. C. Petrantonakis and L. J. Hadjileontiadis, “A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 737–746, 2011.
- [21] B. Stemmer and J. F. Connolly, “The EEG/ERP technologies in linguistic research: An essay on the advantages they offer and a survey of their purveyors,” *Ment. Lex.*, vol. 6, no. 1, pp. 141–170, 2011.
- [22] S. Sur and V. Sinha, “Event-related potential: An overview,” *Ind. Psychiatry J.*, vol. 18, no. 1, p. 70, 2009.

- [23] S. J. Luck, G. F. Woodman, and E. K. Vogel, “Event-related potential studies of attention,” *Trends Cogn. Sci.*, vol. 4, no. 11, pp. 432–440, 2000.
- [24] T. Koenig, K. Kochi, and D. Lehmann, “Event-related electric microstates of the brain differ between words with visual and abstract meaning,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, no. 6, pp. 535–546, 1998.
- [25] E. M. Sokhadze, M. F. Casanova, E. L. Casanova, E. Lamina, D. P. Kelly, and I. Khachidze, “Event-related Potentials (ERP) in Cognitive Neuroscience Research and Applications,” *NeuroRegulation*, vol. 4, no. 1, pp. 14–27, 2017.
- [26] M. Adamo and S. Ferber, “A picture says more than a thousand words: Behavioural and ERP evidence for attentional enhancements due to action affordances,” *Neuropsychologia*, vol. 47, no. 6, pp. 1600–1608, 2009.
- [27] M. Kutas and K. D. Federmeier, “Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP),” *Ssrn*, 2010.
- [28] H. T. Schupp, T. Flaisch, J. Stockburger, and M. Jungho, “Emotion and attention : event-related brain potential studies,” *Brain*, vol. 156, pp. 31–51, 2006.
- [29] J. Kissler, C. Herbert, I. Winkler, and M. Junghofer, “Emotion and attention in visual word processing-An ERP study,” *Biol. Psychol.*, vol. 80, no. 1, pp. 75–83, 2009.
- [30] H. Taha, R. Ibrahim, and A. Khateb, “How does arabic orthographic connectivity modulate brain activity during visual word recognition: An ERP study,” *Brain Topogr.*, vol. 26, no. 2, pp. 292–302, 2013.
- [31] D. E. Dietrich, C. Waller, H. M. Emrich, and T. F. Münte, “Event-related potentials and word recognition memory: differential effect of emotional content,” *Mapp. Cogn. time Sp.*, no. 0, p. 0, 1995.
- [32] S. J. Thomas, S. J. Johnstone, and C. J. Gonsalvez, “Event-related potentials during an emotional Stroop task,” *Int. J. Psychophysiol.*, vol. 63, no. 3, pp. 221–231, 2007.
- [33] G. J. Lewis and T. C. Bates, “Just above Chance: Is It Harder to Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-dependent Signals?,” *Psychologist*,

- vol. 26, no. 3, pp. 194–198, 2013.
- [34] Kondraske, G. V. Neurophysiological measurements. *Biomedical Engineering and Instrumentation*, 138-179, 1986.
- [35] Bronzino, J. D. Principles of electroencephalography. The biomedical engineering handbook, 1995 .
- [36] V. L. Towle *et al.*, “The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, no. 1, pp. 1–6, 1993.
- [37] Da Silv, F. L. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippencott Wlliams & Wilkins , 2005.
- [38] Vespa, Paul M.; Nenov, Val; Nuwer, Marc R., “Continuous EEG Monitoring in the Intensive Care Unit: Early Findings and Clinical Efficacy,” *Journal of Clinical Neurophysiology*. 16 (1): 1–13, 1999.
- [39] T. L. Schultz, “Technical tips: mri compatible eeg electrodes: Advantages, disadvantages, and financial feasibility in a clinical setting,” *Neurodiagn. J.*, pp. 69–81, 2012.
- [40] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, “Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain,” *Rev. Mod. Phys.*, 1993.
- [41] S. O’Regan, S. Faul, and W. Marnane, “Automatic detection of EEG artefacts arising from head movements,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC’10*, 2010.
- [42] K. J. Kieran and J. A. Brunbergz, “Adult claustrophobia, anxiety and sedation in MRI,” *Magn. Reson. Imaging*, 1997.
- [43] J. F. Schenck, “The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds,” *Med. Phys.*, 1996.
- [44] F. Yasuno *et al.*, “The PET radioligand [11C]MePPEP binds reversibly and with high specific signal to cannabinoid CB1 receptors in nonhuman primate brain,”

Neuropsychopharmacology, 2008.

- [45] Mulholland, Thomas , Objective EEG Methods for Studying Covert Shifts of Visual Attention. In McGuigan, F. J.; Schoonover, R. A. (eds.). *The Psychophysiology of Thinking: Studies of Covert Processes*. pp. 109–51, 2012.
- [46] T. Hinterberger, A. Kübler, J. Kaiser, N. Neumann, and N. Birbaumer, “A brain-computer interface (BCI) for the locked-in: Comparison of different EEG classifications for the thought translation device,” *Clin. Neurophysiol.*, 2003.
- [47] Sereno, SC, Rayner, K, Posner, MI , Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport*. 9 (10): 2195–200, 1998.
- [48] I. Feinberg and I. G. Campbell, “Longitudinal sleep EEG trajectories indicate complex patterns of adolescent brain maturation,” *Am. J. Physiol. Integr. Comp. Physiol.*, 2012.
- [49] R. Srinivasan, “Methods to Improve the Spatial Resolution of EEG,” 1999.
- [50] A. Schlögl, M. Slater, and G. Pfurtscheller, “Presence research and EEG.”
- [51] L. Otten and M. Rugg, “Interpreting event-related brain potentials,” *Event-related potentials A methods Handb.*, vol. 72, no. 3, pp. 3–16, 2004.
- [52] M. D. Rugg, M. G. H. Coles, M. G. H. Coles, and M. D. Rugg, “Event-related brain potentials: an introduction,” *Electrophysiol. Mind*, pp. 1–26, 2008.
- [53] G. Kesavaraj and S. Sukumaran, “A study on classification techniques in data mining,” in *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*, 2013.
- [54] K. Q. Weinberger, J. Blitzer, and L. K. Saul, “Distance Metric Learning for Large Margin Nearest Neighbor Classification.”
- [55] Tan, Pang-Ning, M. Steinbach, and V. Kumar , Introduction to data mining." *Library of Congress* , 2006.
- [56] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.

- [57] “Detection Model for Denial-of-Service Attacks using Random Forest and k-Nearest Neighbors,” *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 2, No 5, pp. 1855–1860, 2013.
- [58] S. B. Imandoust and M. Bolandraftar, “Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background,” *J. Eng. Res. Appl. www.ijera.com*, vol. 3, no. December, pp. 605–610, 2013.
- [59] A. Yazdani, T. Ebrahimi, and U. Hoffmann, “Classification of EEG signals using Dempster Shafer theory and a K-nearest neighbor classifier,” in *2009 4th International IEEE/EMBS Conference on Neural Engineering, NER '09*, 2009.
- [60] R. Beale and T. Jackson, “The multilayer perceptron,” *Neural Computing: An Introduction*. 2004.
- [61] S. Haykin, *Neural Networks: A Comprehensive Foundation*. 1994.
- [62] T. Hastie, R. Tibshirani, and J. Friedman, “Springer Series in Statistics The Elements of,” *Math. Intell.*, vol. 27, no. 2, pp. 83–85, 2009.
- [63] D. Anguita, A. Ghio, N. Greco, L. Oneto, and S. Ridella, “Model selection for support vector machines: Advantages and disadvantages of the Machine Learning Theory,” in *Proceedings of the International Joint Conference on Neural Networks*, 2010.
- [64] R. A. Bauder and T. M. Khoshgoftaar, “A Probabilistic Programming Approach for Outlier Detection in Healthcare Claims,” in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2016, pp. 347–354.
- [65] M. K. Stern, J. E. Beck, and B. P. Woolf, “Naive Bayes classifiers for user modeling,” *Cent. Knowl. Commun. ...*, pp. 1–10, 1999.
- [66] T. R. Patil and M. S. S. Sherekar, “Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification,” *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, 2013.
- [67] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *Proc. 14th Int. Jt. Conf. Artif. Intell. - Vol. 2*, vol. 2, no. 0, pp. 1137–1143, 1995.

[68] K. M. Ting, “Confusion Matrix,” *Encycl. Mach. Learn. Data Min.*, pp. 260–260, 2017.

[69] K. M. Ting, “Confusion Matrix,” *Encycl. Mach. Learn. Data Min.*, pp. 260–260, 2017.

