Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

COLLEGE OF ENGINEERING
ELECTRICAL & COMPUTER ENGINEERING DEPARTMENT

ECE CCEE595B Summative Learning Project 2 (SLP2)

Spring 2021

Sentiment Mining and Analysis

Group & Disciplines

Ahmad Cheble — *20170149 CCE* ***(Computer and Communication Engineering)***

Hassan Khamis — *20170147 BE* ***(Biomedical Engineering)***

Advisor: *Professor* ***Rached Zantout****, Electrical and Computer Engineering department at RHU*
zantoutrn@rhu.edu.lb

Co-advisor: *Doctor* ***Samir Berjawi****, Electrical and Computer Engineering department at RHU*
berjaouisw@rhu.edu.lb

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

1- Overview:

This project will apply Artificial Intelligence, Machine Learning and Data mining, Computer vision, Deep learning, and natural language processing concepts to be able to detect the sentiment of a person. Concerning the text classification of tweets phase, the system will prepare a dataset, visualize it, preprocess, and clean it, using emoticon, acronym, stop words, positive and negative dictionaries, by replacing them with a certain tag. In addition, this system will replace URLS, usernames, retweets, negations with their corresponding tag as well. It will reduce all letters to lower case and will decode HTML entities and remove Unicode characters. Then, feature selection and extraction, tagging, tokenization, stemming, lemmatization and other procedures will be performed. Moreover, the system will then use machine learning classifiers like Naïve bayes, SVM, Logistic Regression and many others, to detect the sentiment. Training, folding, validation and visualizing results will be performed. After choosing the best classifier according to the prediction metrics. The classifier will be used for testing. Retrieving tweets posted by a certain individual will be performed, and the testing shall start. Deep learning can be also used to detect the sentiment, and after comparing both deep learning and machine learning, the best approach will be the applicable one. On the other hand, concerning the facial expressions detection, sentiment analysis will be performed based on recorded interviews, sessions, and meetings, using a new Deep learning library launched by Facebook, called Deep face and other video editing libraries and functions. This library will be also used to detect the sentiment based on social media posts, it will simply give the dominant sentiment to pictures after locating the face and detecting it. This will require computer vision and deep learning. Concerning, vital signs, the system will try to detect the sentiment of an individual using sensors, to acquire the data during an interview for example, and the data will be used in machine learning algorithms, which will give a prediction to the sentiment of an individual based on vital signs, that could be used to validation purposes as well. Finally, this system can be used in many domains, whether to analyze sentiments for certain desires, or to enhance negative sentiments and make them positive ones, and finally, to solve real life problems and issues.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

2- Acknowledgments:

This project was done as our final year project at Rafik Hariri University. Supervised by Prof. Rached Zantout and co-advisor Dr. Samir Berjawi. After finishing up all the research and literature review needed for this topic during the SLP1, we had 3-4 months to finish everything, including the implementation and testing of the whole system. Every couple of weeks, meetings were done with our advisors, to manage our obstacles, see our progress, and plan for the upcoming tasks.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Table of Contents

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Table of Figures

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

3- Introduction

3.1- Motivations and interests

As a Computer and Communications Engineering student at RHU, I was extremely interested in Artificial Intelligence, Programming, mainly Python, real-life applications, and Technology. However, personally I was always interested in Social Media, and the latest trending topics and discussions worldwide. So, this combination motivated me to start working on this project. I believe that we came to this world to make a change, make it easier for future generations by solving real life problems using our expertise and knowledge. We all have fingerprints that no one else has, why cannot we leave an imprint no one else did, and I am planning to leave my imprint through this project. During my first 3 years at RHU, I attended several hands-on Deep Learning seminars and workshops presented by NVIDIA Deep learning institute and earned certificates in this domain, I also attended several python workshops and competitions, not to forget that I participated at two local robotics, programming, and AI related competitions in Lebanon mainly FLL and NERD competitions, and one global international competition (FLL open international) that was also held in Lebanon in 2019. All these boosted my interest and gave me a kind of clear image of what I like to do, and what I want to be an expert in. So, I took Natural Language Processing, Machine Learning and Data Mining, and Artificial Intelligence courses at RHU, and after finishing these courses with very good grades, I have not had enough of them, I wanted to explore more about these tracks. All these reasons made me pick this topic as my Final year project at RHU. I also applied to different master's degrees abroad in these tracks as well and I started to receive acceptances. Concerning the project, I used Natural Language Processing, Data Science, Machine learning and Data Mining for sentiment detection of tweets with text classifications. In addition to that, I used latest Deep learning libraries and Computer Vision for facial detection and analyzing images from social media posts or pictures, that will provide us with sentiment detection and some other purposes as well. I also used Computer Vision, Video editing libraries, and Arrays for sentiment detection of interview videos or any other kind of videos as well. More information will be presented later in this report.

As a Biomedical Engineering student at RHU, I was interested in exploring the metaphysics of data and what can be implemented from it. This takes a lot of deep analysis and understanding of Artificial Intelligence (AI). With good planning and advising, the combination of analysis and AI can make a powerful duo and lead to an attractive and interesting project. Leaving a positive mark is something everyone looks to do, and that is what we want to do in this project, to do something that interests us and at the same time to do something special and unique that has not been done before. The whole world right now is booming with artificial intelligence in all the fields, so I was interested in grasping the idea of it and the way it works. It all starts by triggering the interest and followed by a healthy surrounding which was provided by the advisors, nothing would be impossible. This interest was first initiated from a project that I did which requires intensive use of AI; having worked on VGRF fatigue detection for marathon runners which uses a lot of data analysis, machine learning, data classification and MATLAB skills increased my interests in the AI topic. Also, I was always fascinated by the psychology of humans and what really influences them. That is why I wanted to relate AI in a topic that I am personally interested in which is to detect the sentiment of people. More details will be presented regarding this topic.

3.2- Purposes

After the Covid-19 crisis, the world we are living in has changed, governments are facing hard economical situations, people are facing serious psychological problems (anger, anxiety, depression, and stress). Everyone is locked inside their house. Companies and universities are losing a lot, hiring became harder, and accepting students became difficult. In addition, relationships are getting worst, crimes of all kinds are increasing and finally, politicians, role models and famous people around the world, are posting unrelated tweets and miss lead information. On the other hand, social media and technology are rising tremendously, everyone is on social media more than before, especially in the current crisis. As a result, a solution to all these problems would be a system that can use Social media to detect the

sentiment of an individual and provide solutions to enhance the individual's mood in a positive way. Not to forget that, as mentioned before, this project can be used in several domains for example: Health care (making a therapy for patients suffering from anxiety for example), Crimes (during investigations with suspects in serious murders and other cases for example), Relationships between couples (during meetings with a psychiatric that would try to identify the problems and solve them) , Interviews (while interviewing an applicant for a certain job) and last but not least, it can be also used during meetings with politicians, role models and famous people. All these tracks are not the only ones. Sentiment analysis and detection is very important, and it will stay important in the future, because a human being's feelings or emotions cannot disappear and cannot be substituted. In fact, if we want to speak out of an AI perspective, we can say that it might be one of the only things that robots cannot have, and cannot be considered as a substitute for, in the future.

3.3- Sources

Since we want to share our experience and knowledge we acquired to succeed in this project, we will show all the sources of information that will help others when trying whatever we have done in this project.

3.3.1 For Sentiment Analysis of tweets, social media posts and pictures, videos and interviews, these sources are important to know and have experience with, to succeed in this project.

Github Link will be provided for this part of the code to help you succeed in such project **[0]**

Jupyter Notebooks (used in this project) are free open documents based on JSON. They are used to combine software code, computational outputs, and multimedia resources in a single document. It is widely used for AI applications including data cleaning, machine learning, computer vision, deep learning, NLP, data mining and several other tasks as well. We used it to implement Python programming language. More information is presented in the reference: **[1]**

Python (Version 3 was used in this project) is an interpreted, object-oriented, high-level

programming language with dynamic semantics. It is simple and easy to learn. It is widely used in all kinds of applications. This project includes knowledge with functions, for loops, if else, while, arrays, main algorithms and variables, operations, concepts, and everything presented in this report. More information is presented in the reference: **[2]**

Natural Language Processing is a subfield of Artificial intelligence, computer science and linguistics. It is mostly related to human to computer interaction. It is basically how to program computers, to process and analyze large amounts of data. It used in a variety of application related to AI and machine learning. It widely used in text classification and processing as well. Knowledge in this field is required. Different topics included in this domain mainly: **Regular Expressions, Stemming, Lemmatization, N-grams,** and many other tracks. These are strongly required for an individual to succeed in such project. More information is presented in the reference: **[3] [4]**

Machine learning is the study of computer algorithms that improve through experience and using data. It is considered as a subset of Artificial intelligence. It gives computer programs the ability to access data and use it to learn for themselves. The different steps that this topic includes mainly: **Data Preparation, Data Preprocessing, Feature extraction, Feature Selection, Training the data, Folding, Cross validations, Evaluation, choosing best model, Testing,** and many other stuffs, are strongly required for an individual to succeed in this project. More information is presented in the reference:**[5][6][7]**

Data Mining is the process of extracting and discovering patterns in large datasets. It includes methods and concepts that meet with the machine learning fundamentals. Knowledge in this field is a strong backup to succeed in this project. More information is presented in the reference:**[8][9][10]**

Deep Learning is a subset of machine learning in artificial intelligence, that has neural networks capable of learning unsupervised from data. More information is presented in the reference:**[11][12]**

Computer Vision is a field of Artificial Intelligence that trains computers to understand the visual world. It gives computers the ability to use images, videos, and deep learning models. It is used for identification, classification, detection, and many other fields. Knowledge in this domain is needed as well

to succeed in this project. More information is presented in the reference: **[13][14][15]**

DeepFace library which is a deep learning facial recognition system created by a research group at Facebook, it was released in December 2020, and its considered one of the latest most trending libraries in the field of image processing, detection, verification, and analysis. This library can be installed in Jupyter notebook, and it identifies human faces in digital images. It employs a 9-layer neural network with over 120 million connection weights, in addition, this library was trained on four million images upload by Facebook users. It is stated that Deepface reaches approximately 97.35% plus or minus 0.25% on labeled faces in LFW dataset where human beings have 97.53%. So, DeepFace can sometimes be more successful than the human beings. More information is presented in the reference:**[16][17][18]**

Scikit-learn and can be known as sklearn. It is a free software and machine learning based library for python programming language. It is used for various applications including: Classification, Clustering, Regression. It also used for several algorithms like: Naïve Bayes, SVM, Random Forest, Decision Trees, and many others. This library is widely used in the field of Artificial intelligence and Machine Learning and is considered the most trending libraries in applications related to these fields. More information is presented in the reference:**[19][20]**

Scikit-learn tools like folding, metrics (confusion matrix, f1_score, precision score, accuracy score, recall score) in addition to, the classifiers mainly used here are (Multinomial Naïve Bayes, KNeighbor, SGD, Logistic Regression, SVM, Decision Tree), also, the feature extraction tool which is the count vectorizer function. Those are a must for an individual to succeed in this project, those are a major part of the text classification, and are considered extremely important. More information is presented in the reference:**[21] [22] [23] [24] [25] [26] [27] [28] [29]**

Tweepy is a library presented in python, it is used so that users can access Twitter API. It is used for automation and creating twitter bots. It is also used in a variety of applications used in text classification and processing of twitter data. More information is presented in the reference:**[30]**

RE which is known as regular expressions. It is a group or a sequence of specific characters, these

characters specify or represent a certain search pattern. It widely used in the field of text classification and preprocessing, text processing and analyzing. It is one of the most trending topics lately used in the field of Natural Language Processing. More information is presented in the reference:**[31][32]**

NLTK is a python library commonly used in Natural Language Processing applications. It is used for classification, tokenization, stemming, tagging, parsing and several other tasks and fields. It is also widely used in the preprocessing stages of text classification applications. More information is presented in the reference:**[33][34]**

Tkinter which is a python blinding to Tk GUI toolkit. It is a python library used to give the user access with graphical user interface dialogs, that will facilitate to the user the implementation and testing phases and will save time. This library is included with standard Linux, Microsoft windows and Mac OS X installs of python. More information is presented in the reference:**[35][36]**

OpenCV and CV2 libraries are mainly aimed towards real-time computer vision applications in python. They are open-source libraries, and they are free to use. Mainly used in the field of image processing, video captures, analysis, and detection. More information is presented in the reference:**[37][38]**

Pandas Library is a software library, used in python programming language. It is used for data manipulation and analysis. It offers many services, mainly data structures and operations for numerical tables and time series. Its free to use and it is an open source. More information is presented in the reference: **[39][40][41]**

Matplotlib is a plotting library used in Python programming language. It delivers object-oriented API in terms of embedding and plotting into applications with the use of GUI. It helps users visualize their data in a neat and organized manner. It is widely used for data visualization and presentation. More information is presented in the reference:**[42][43]**

Pickle is a library for python programming language. It is used for implementing binary protocols

for serializing and de-serializing objects in python. It is also used to saving and loading models used for classification applications. More information is presented in the reference:**[44]**

Access tokens are tokens needed for a user to receive access to twitter data. Certain questions will be asked, and an application must be filled online, and the user shall receive access with the necessary information and tokens needed to write inside the python programming code, so that the user gets full access to twitter data. More information is presented in the reference:**[45]**

NumPy is a library used in python programing language. It is widely used in the field of multi-dimensional arrays and matrices; it is also used in large collection of high-level mathematical functions to operate these arrays. More information is presented in the reference:**[46][47]**

FFmpeg is a free software and an open-source library used in python for handling videos, audios and other multimedia files and streams. More information is presented in the reference:**[48][49][50]**

MoviePy is a module presented in python mainly for video editing tasks. This module can be also used for cuts, title insertions and many other tasks as well. It widely used in the field of video editing and operations in python applications. More information is presented in the reference:**[51]**

wordcloud is a library used for data visualization techniques used for python programming language. It represents data according to frequency or importance, and widely used in text classification applications and data science projects. More information is presented in the reference:**[52]**

html is a python library used for decoding HTML entities. It is widely used in the field of text classifications and preprocessing. More information is presented in the reference:**[53]**

Collections is a python library that has some datatypes and containers, which provide the user alternatives for python built-in containers. It is widely used in all applications. More information is presented in the reference:**[54]**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Textblob is a python library used for text processing. It can be used in tagging, translation, classification, and sentiment analysis tasks. It is a very well-known library in the field of Natural Language Processing. More information is presented in the reference:**[55][56]**

Google Collab is a product from Google research. It allows anybody to write and execute python codes through the browser. It is used for various applications including deep learning, machine learning, AI, and computer vision. We did not use it a lot in our project, but it can be a substitute for jupyter notebook if the user decides to, but the code shall be different sometimes, and may require some knowledge about it. More information is presented in the reference:**[57][58]**

os is a python library that provides functions for interacting with the operating system. It provides a portable way of using the operating system-dependent functionality. More information is presented in the reference:**[59]**

time is a python library that provides many ways of representing the time in python code implementation, whether as objects, numbers, or strings. It provides functionality like waiting code execution and measuring efficiency of the code. More information is presented in the reference:**[60]**

csv is a python library used for reading and writing csv files in python programming language. It is widely used for applications that include dealing with dataset and csv files. More information is presented in the reference: **[61]**

3.3.1.1- Articles and reports:

Twitter Emotion Analysis is a report done by Mr. Marc Lamberti where he made text classification for twitter data and he used several datasets for negations, acronyms, positive and negative words, and smileys. His best improvements and last result were an F1 score of 79.5% using bigrams and unigrams only along with multinomial Naïve Bayes. Reading his article and code would help you proceed with such project. More information is presented in the reference:**[62] [63] [64] [65] [66] [67]**

Emotion and Sentiment Analysis from Twitter Text is a master thesis from university of Calgary published in 2018. It shows all preprocessing techniques used for sentiment analysis of tweets in addition, it shows statistics about deeper information in the twitter data, like retweets, replies and followings. Reading this thesis would give you some ideas to work on in this project that would improve the system, and it is important to read before making such project related to twitter. More information is presented in the reference.**[68] [69] [70] [71] [72] [73] [74] [75]**

Sentiment Analysis of Twitter Data: A survey of Techniques is a journal published by the CE Department of University of Pune in India. It used Naïve Bayes, SVM, Maximum entropy algorithms with different N-Grams, and reached an accuracy of 77.73% with SVM with bigrams only. Reading this paper is very beneficial for anyone who wants to work in such project or domain. More information is presented in the reference:**[76] [77] [78] [79]**

Emotion Detection of Tweets using the AIT-2018 Dataset is a text classification paper which was published in 2019. It clearly states the preprocessing phases needed for text classification and sentiment analysis of tweets. Reading this article will help a lot in terms of data cleaning and preprocessing. More information is presented in the reference:**[80] [81] [82] [83][84][85]**

Detecting Emotions in Twitter Messages is a paper published by Worcester Polytechnic Institute CS Department in 2014. It clearly states preprocessing techniques for twitter text classification, in addition to that, it also explains feature selection process and N-gram features as well. These information are beneficial and good to read. More information is presented in the reference:**[86][87]**

Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks is a paper published by the Department of Electronics and Communication Engineering at the National Institute of Technology in Mangalore India. It shows the main flowchart of the methodology required to follow, to do the emotion classification successfully. It was helpful at start, but after the Deep face library got published, it saved a lot of work for us, and detection and classification were a piece of cake for us. Reading it is important but not necessary. More information is presented in the reference:**[88] [89]**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

3.3.2 For Sentiment Analysis of vital signs and different rates and body signals with signal processing these sources are important to know and have experience with, to succeed in this project.

Matlab Software, MathWorks created MATLAB, a multi-paradigm programming language and numeric computing environment. Matrix manipulations, function and data plotting, algorithm execution, user interface creation, and interfacing with programs written in other languages are all possible with MATLAB. **[93]**

Arduino IDE, the Arduino Integrated Development Environment (IDE) is a cross-platform program written in C and C++ functions. It is used to write and upload programs to Arduino-compatible boards, as well as other manufacturer development boards with the support of third-party cores. **[94]**

VScode, Microsoft's Visual Studio Code is a freeware source-code editor for Windows, Linux, and macOS. Support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git are just a few of the features. **[95]**

GitHub, Inc. is an Internet hosting company that specializes in Git-based software creation and version control. It includes Git's distributed version control and source code management capabilities, as well as its own. **[96]**

Classification Learner, Classification Learner allows you to interactively explore the files, pick attributes, define validation schemes, train models, and evaluate outcomes, among other supervised learning activities. **[97]**

MAX30102 sensor, The MAX30102 is a biosensor package with integrated pulse oximetry and heart rate control. Internal LEDs, photodetectors, optical components, and low-noise electronics with ambient light rejection are all part of the package. The MAX30102 is a full machine solution that makes design-in for smartphone and wearable devices even easier. **[98]**

Arduino UNO, The Arduino Uno is an open-source microcontroller board designed by Arduino.cc and built on the Microchip ATmega328P microcontroller. The board has optical and analog input/output pins that can be used to connect to various expansion boards and circuits. **[99]**

SparkFun Max30105 library, is a library provided by saprkfun and works on different types Max301x sensors . **[100]**

Max30102 library, is an integrated Arduino IDE library for the Max30102 sensor **[101]**

Electronicclinic, is the website that clarifies the working procedure of the sensor. **[109]**

Signal Processing knowledge, this indicates how to implement signals and be able to read certain data from signals. **[110] [111] [112][113][114][115][116][117]**

3.3.2.1- Articles and reports:

Human Vital Signs Detection Methods and Potential Using Radars: A Review, is an article made by Kebe et. al which dugs deep in the methods of detecting human vital signs. It mostly relies on 12 lead ECG based system which is the traditional one. Or the continuous wave (CW) radar that can detect vital signs wirelessly. **[103]**

Human Emotion Recognition: Review of Sensors and Methods, an article that focuses on the

emotion detection using data collected from sensors. Features from the heart signal like the HRV were collected to facilitate/increase the accuracy of the detection. **[104]**

Detection of Stress Levels from Bio signals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine is a work that relates stress levels to detected biosignal data using extreme machine learning. **[105]**

Wearable Emotion Recognition Using Heart Rate Data from a Smart Bracelet is an article done by shu et. al that detects the mood of people by measuring their heart rate. In this work, database was collected while watching a stimulant video for each mood. **[106]**

Emotion Recognition based on Heart Rate and Skin Conductance, a study designed to relate the ECG, EDA and EMG to different types of emotions like joy, anger, disgust, sadness. **[107]**

3.4- Manuscript Structure:

We will be showing the sentiment analysis of images and social media posts, interview videos and meetings, vital signs and finally, tweets posted by twitter users on Twitter application. All steps will be shown clearly with the complete procedure required.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

4- Project Scope statement:

A survey (online survey) was built to summarize interaction with possible customers and end-users, to explain the problem statement of clients, consumers, or end-users. The objectives will be shown as well in the survey. In addition to that, constraints, complications, and deliverables will be shown and listed as well.

4.1- Survey Analysis:

Questions:

Are you a? *

Student

Professor / Doctor / Lecturer

Employee

Is this the first time you hear of Sentiment Analysis? *

Yes

No

Are you a moody person? *

Yes

No

Does your mood affect your relationship with surroundings? *

Yes

No

Do you often tell your surroundings about your mood? *

Yes

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

No

Do you think People with a positive mood are more productive than employees with negative mood? *

Yes

No

How often do you get negative feelings? *

Not much

Recently

All time

Do you think someone's posts or pictures on social media can reflect his/her mood? *

Yes

No

Do you think someone's tweets or captions can reflect his /her mood? *

Yes

No

Do you think vital signs can reflect someone's mood? *

Yes

No

Do you think having a product that tells you the mood of a certain person is a useful system to be implemented? *

Yes

No

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

In your opinion who can benefit a lot from such a Product?

Your answer

In your opinion where do you think the product can be used the most?

Your answer

Do you think the device should be wearable or not? *

Yes

No

Can you rate the importance of the system? *

Above 80%

Between 50% and 80%

Below 50%

Is it good to have multiple moods for the product to detect? *

Yes

No

Do you think the product should give a live detection? *

Yes

No

How much would you pay for such a product? *

Not more than 100$

Between 100$ and 500$

above 500$

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

which moods do you think this product must detect? *

Fear

Anger

Sadness

Happy

Surprise

Disgust

worried

anticipation

Other:

Due to the Covid-19 Situation in Lebanon, the survey was done online, and it reached up to 105 individuals, half of them were students who are enrolled in the Department of electrical and computer engineering, the other half are doctors, lecturers, professors, employees that work in such domain as well.

After analyzing the results:

- Half of the people know about sentiment analysis another half do not
- Most of the people are moody that gives us an advantage about our domain
- Most of the people said that their mood affects their relationship with surroundings and usually they do not tell their mood to people around them which is considered another advantage
- the majority said that people who has positive mood are more productive than people who are not which gives an advantage for us to use the device
- The majority said that they often get negative feelings
- A huge voting went yes to the question concerning if social media, caption, tweets, texts, vital signs are important factors in detecting the mood, which makes our main project points validated and extremely important
- Most of the votes said the device should not we wearable
- Most of the votes rated the importance of our system between 50% and 80% 66 votes, where 22 votes said its above 80%

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

- a lot of votes said that the device is extremely important to be implemented 98 votes
- half of the votes said that the device should not give live detection
- half of the votes said that the device price should be between 100$ and 500$ (56 votes)
- The most voted moods that must be detected were (fear, anger, sad, happy) which we will give a priority in our work
- A very interesting fact that a lot of votes were unique about where and who should use the device here are some examples of what the votes were:

where:

- Hiring employees
- Interviews with politicians
- Investigations
- Hospitals
- Schools
- Universities
- Relationships

Who:

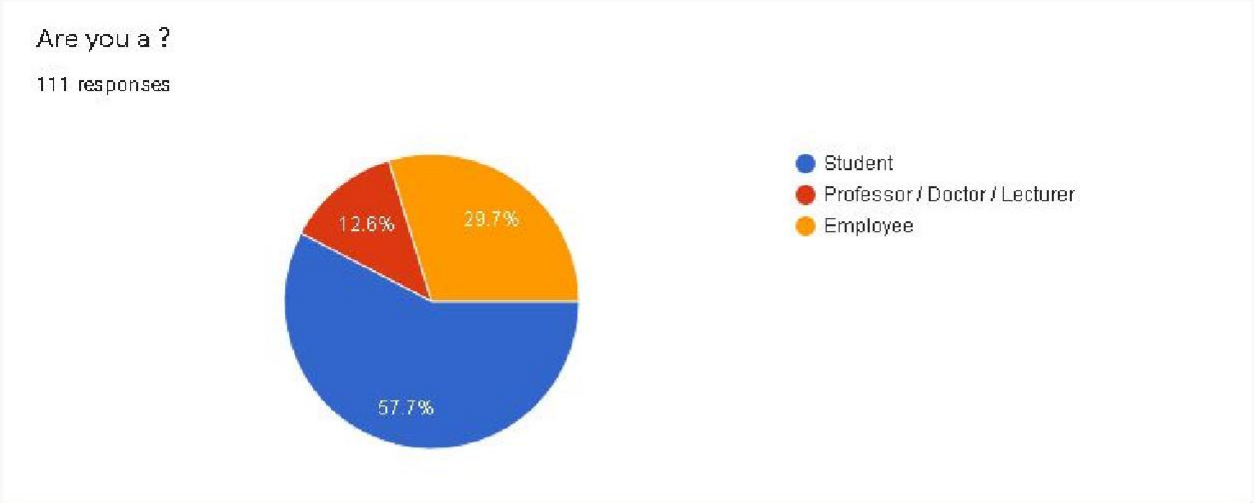
- Companies
- Investigators
- Lecturers
- Interviewers
- Psychiatrists
- Relationship partners

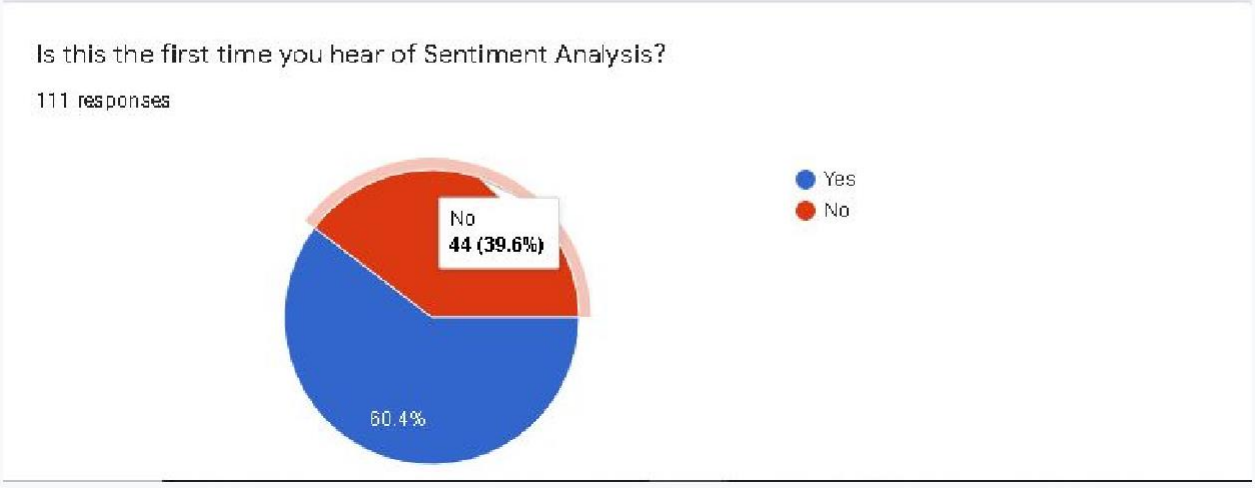
All these answers give us a lot of confidence and trust in our device, the variety of the answers in these two questions which were not mandatory questions to finish the survey, proves that such device will have a variety of fields of usage, thus having a huge success.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

4.1.1- Figures of Survey Results



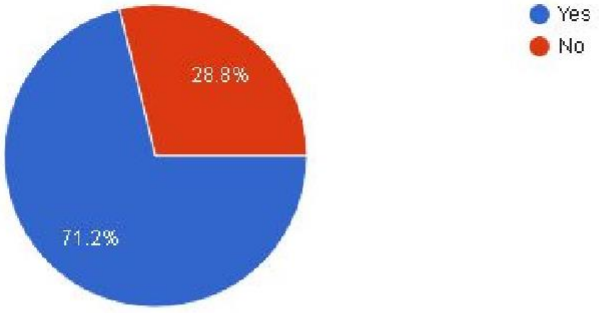


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

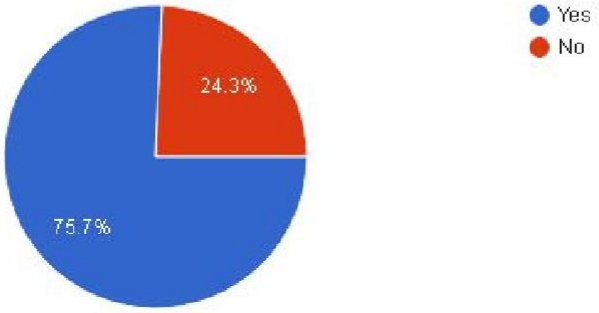
Are you a moody person ?

111 responses



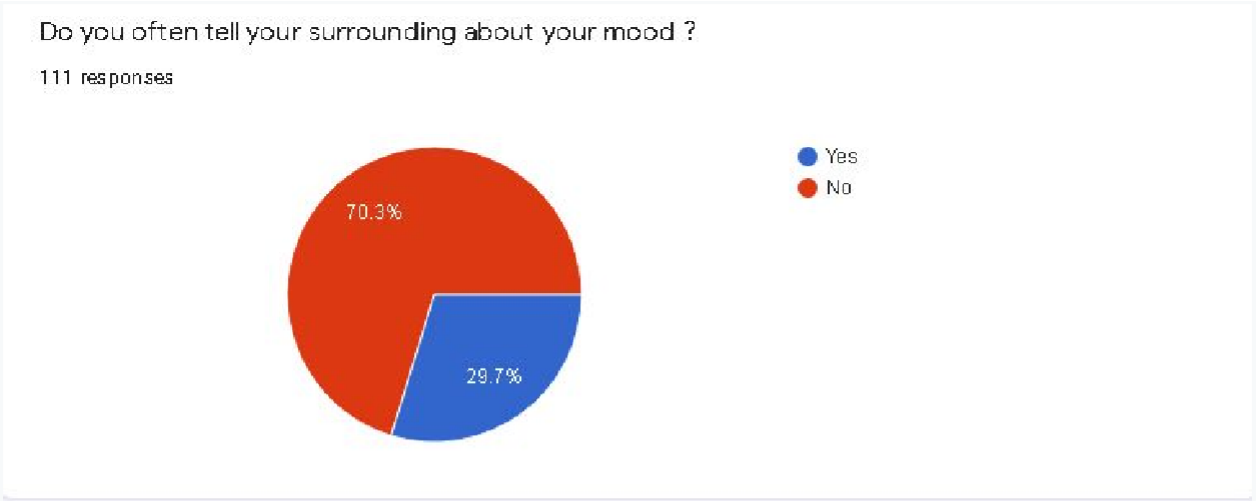
Does your mood affect your relation with surroundings ?

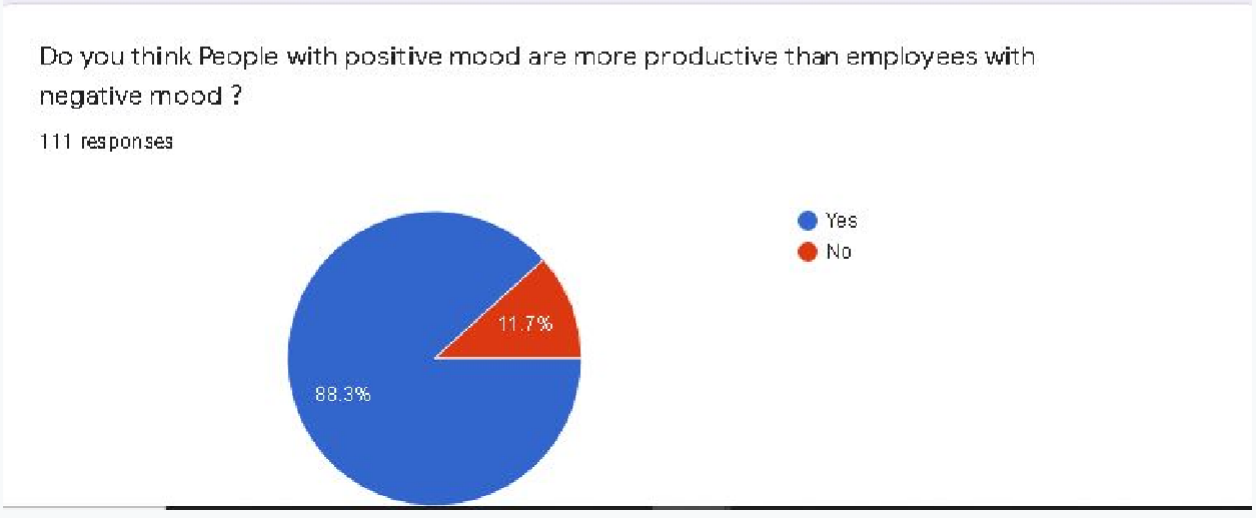
111 responses



Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com



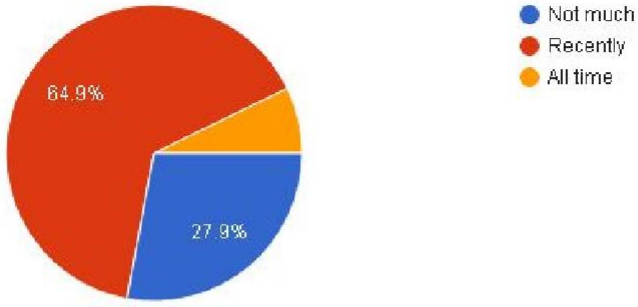


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

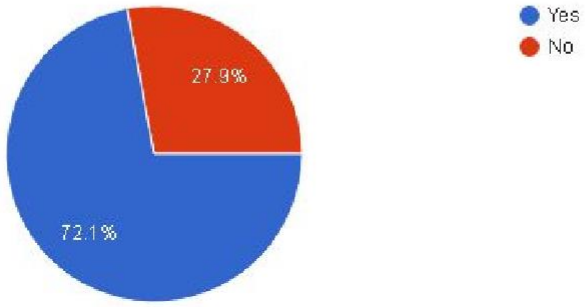
How often do you get negative feelings ?

111 responses



Do you think someone's posts or pictures on social media can reflect his/her mood ?

111 responses

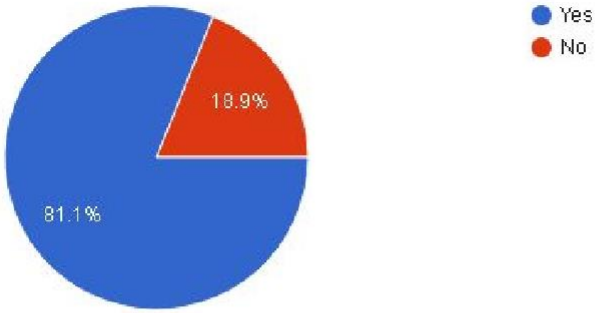


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

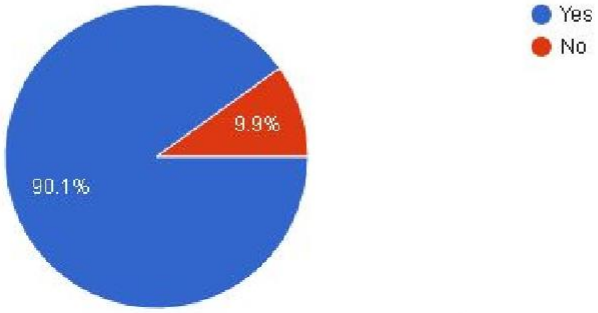
Do you think someone's tweets or captions can reflect his /her mood ?

111 responses



Do you think vital signs can reflect someone's mood ?

111 responses

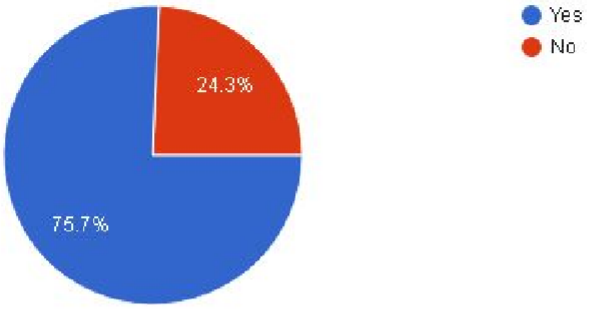


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Do you think having a product that tells you the mood of a certain person is a useful system to be implemented ?

111 responses



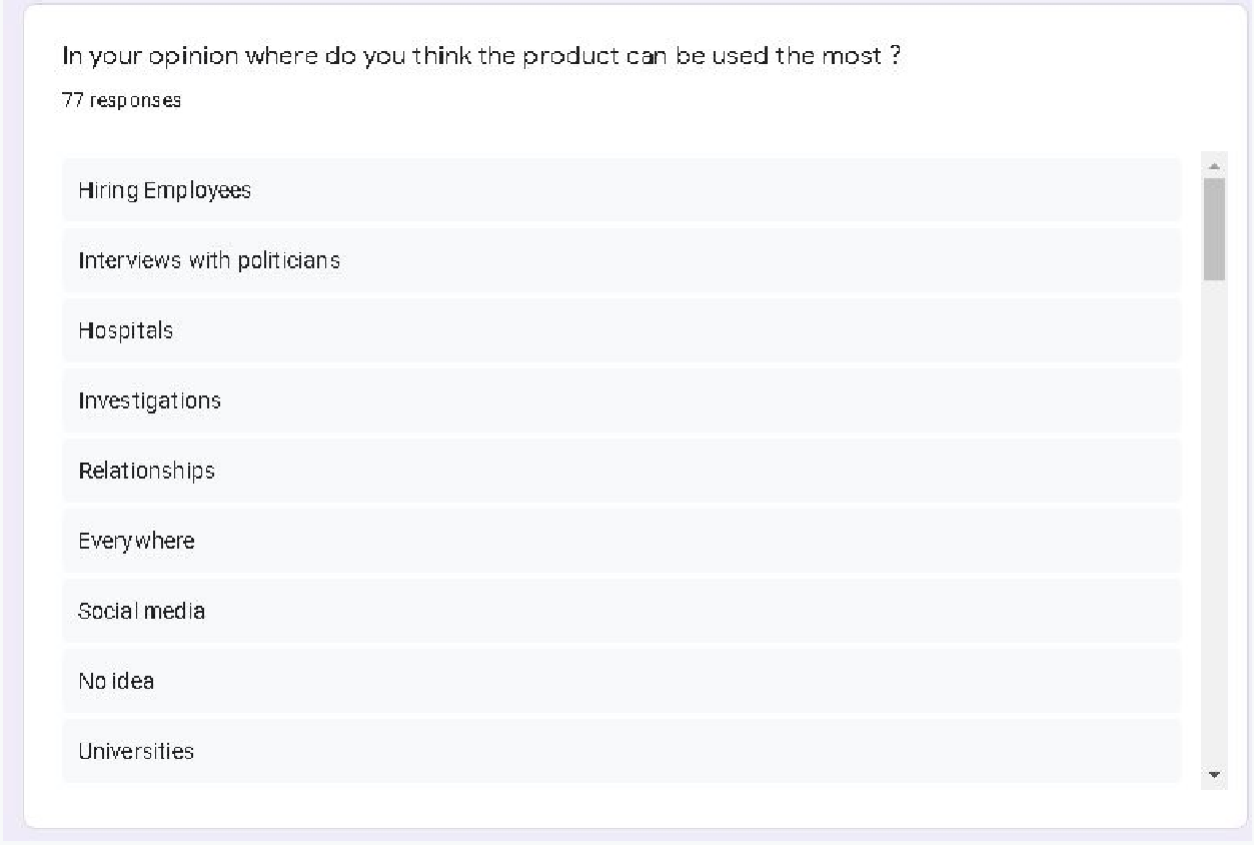
In your opinion who can benefit a lot from such Product ?

85 responses

- Companies
- Investigators
- Lecturers
- Interviewers

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

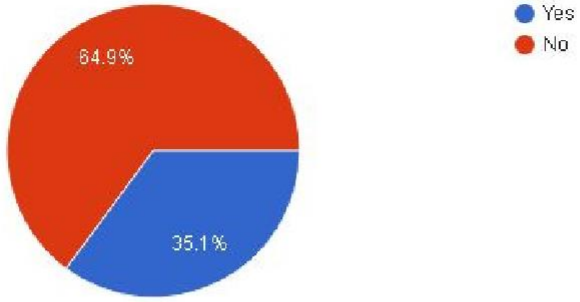


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

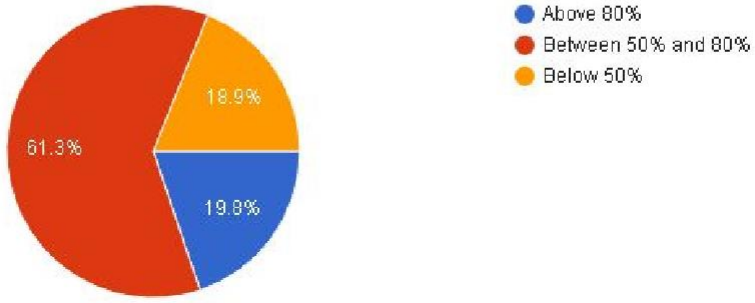
do you think the device should we wearable or not?

111 responses



Can you rate the importance of the system?

111 responses

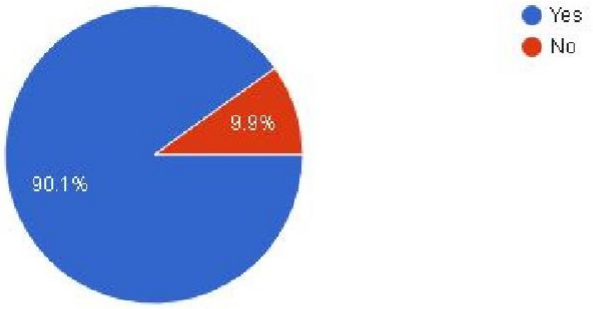


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

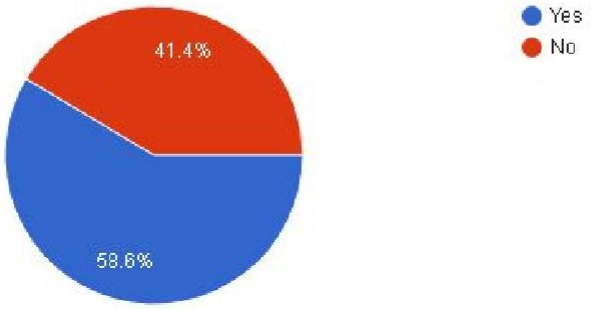
is it good to have multiple moods for the product to detect?

111 responses



Do you think the product should give a live detection ?

111 responses

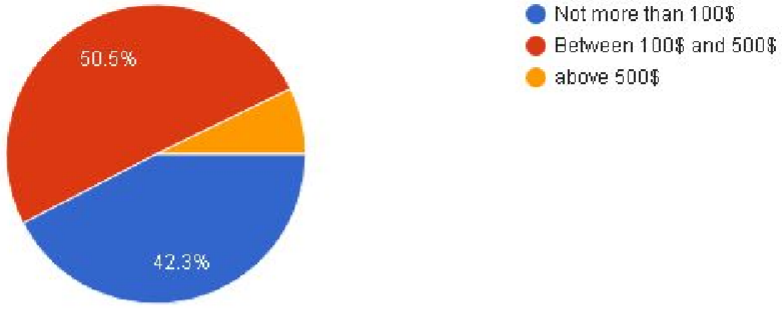


Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

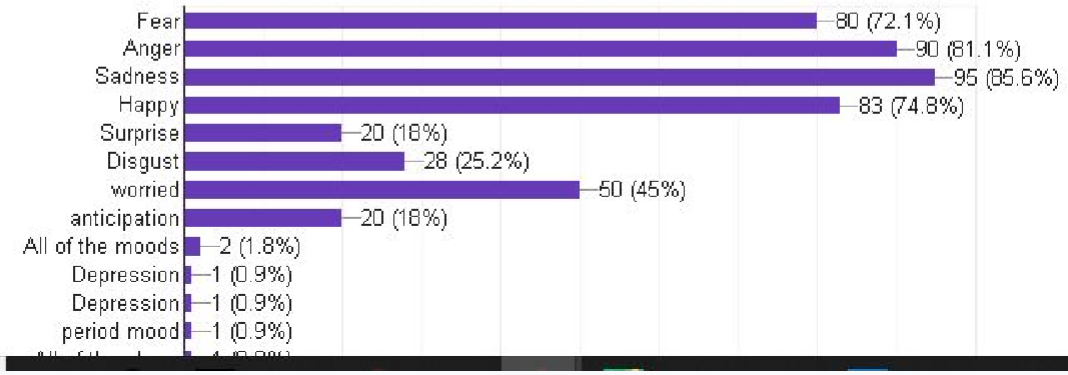
How much would you pay for such product?

111 responses



which moods do you think this product must detect ?

111 responses



Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

4.2- Feasibility Study:

4.2.1- Constraints

- Financial Costs
- Time period to finish the project implementation
- Quality, expectations, and results

4.2.2- Complications

- Motion artifacts that may disrupt the measurement of biological signs
- Evaluation of classifiers and choosing optimal ones
- Metrics to identify time slots for sentiment analysis of interviews (questions)

4.2.3- List of Deliverables

We are expected to perform sentiment analysis of tweets, social media posts and images, vital signs and finally, interview videos.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

5- Data Preparation

5.1 Sentiment Analysis of Twitter Data:

Every classification task needs a balanced dataset with equal number of labels, and in this case, we need the number of positive tweets, to be approximately equal to the number of negative tweets, and a large dataset is needed as well. No to forget that, the increase in the amount of data we have, leads to an increase in the training data and thus having a better accuracy as well.

After searching for datasets, I found one that contains 1578612 English tweets, these tweets are provided from two different sources, Kaggle and from Sentiment140. This dataset has 4 columns as shown in figure 1, but we only case about column "SentimentText" and the "Sentiment" column. However, I called it as the second dataset in my code, because my first dataset is a sample of this dataset. I used this sample to run more algorithms for classification. Further information will be provided later in this report.

```
In [4]: #Display the first 10 rows of the second dataset
        data2.head(10)
```

Out[4]:

	ItemID	Sentiment	SentimentSource	SentimentText
0	1	0	Sentiment140	is so sad for my APL friend.....
1	2	0	Sentiment140	I missed the New Moon trailer...
2	3	1	Sentiment140	omg its already 7:30 :O
3	4	0	Sentiment140	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
4	5	0	Sentiment140	i think mi bf is cheating on me!!! T_T
5	6	0	Sentiment140	or i just worry too much?
6	7	1	Sentiment140	Juuuuuuuuuuuuuuuussssst Chillin!!
7	8	0	Sentiment140	Sunny Again Work Tomorrow :-| TV Tonight
8	9	1	Sentiment140	handed in my uniform today . i miss you already
9	10	1	Sentiment140	hmmm.... i wonder how she my number @-)

Figure 1: Sample Tweets with their corresponding Sentiments whether its negative (0) or positive (1)

Since we only care about the Sentiment Text and the Sentiment columns, I dropped the unnecessary columns and this are the dataset after dropping.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

```
In [5]: # Drop unneeded columns in the Second dataset
        data2=data2.drop(['ItemID','SentimentSource'], axis = 1)
        data2
```

Out[5]:

	Sentiment	SentimentText
0	0	is so sad for my APL friend.....
1	0	I missed the New Moon trailer...
2	1	omg its already 7:30 :O
3	0	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
4	0	i think mi bf is cheating on me!!! T_T
...	...	...
1578607	1	Zzzzzz.... Finally! Night tweeters!
1578608	1	Zzzzzzz, sleep well people
1578609	0	ZzzZzZzzzZ... wait no I have homework.
1578610	0	ZzZzzzZZZZzzz meh, what am I doing up again?
1578611	0	Zzzzzzzzzzzzzzzzzz, I wish

1578612 rows × 2 columns

Figure 2: The dataset after dropping unneeded columns.

In the *Fig1* we can notice the uncleaned tweets that are shown, which gives attention towards the preprocessing phase.

- The **acronyms** are highly presented for example “APL”, we do not know the meaning of this acronym maybe it means apple, or some other significance, this should be dealt with.

- The **repeated characters** like “mmmmmm” “oooooook”. The repition of character in a word increase the negative impact on our system and classification procedure.

- The **Emoticons** for example: “:)” “:O” and many others are highly important as well.

- The **nouns** for example:TV

In addition to these,

- Some indications of emotions presented by users like *cries* and *laughs*.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

- Presence of negations like “wont” “don’t” “cannot”, these negations need to be handled, because they are highly important.

The grammar structure of the tweets is extremely important in our case. As noticed before, it is not an easy task to deal with these issues in language, since we want to analyze tweets posted by twitter users on the application online since users do not give any importance towards the grammar behind the stuff and texts being posted online. That is a huge struggle in text classification. Let us visualize the big dataset positive and negative values count.

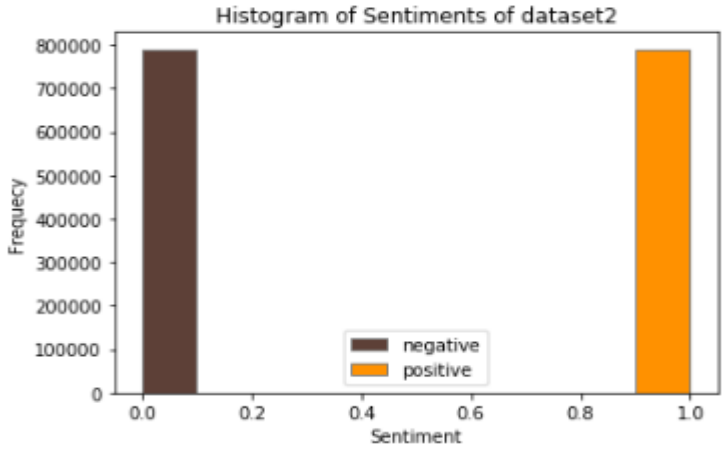


Figure 3: A histogram of sentiments to the frequency of tweets in dataset2

Concerning the first dataset which is a sample of the second dataset. Its manually set up to have equal number of positive and negative tweets. Let us visualize the first dataset as well.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

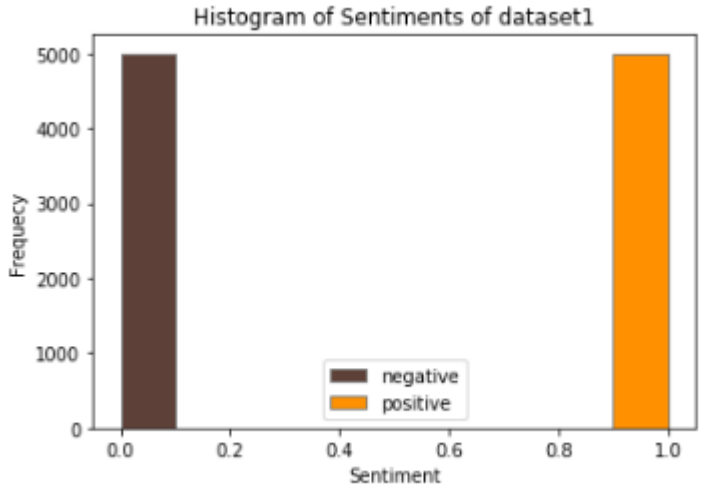


Figure 4: Histogram of sentiments to the frequency of tweets in dataset1

Let us check the exact number of tweets corresponding to each sentiment in both datasets.

```
In [11]: # Dataset 1
         data.Sentiment.value_counts()
         # Count of tweets corresponding to the positive and negative sentiments

Out[11]: 1    5000
         0    5000
         Name: Sentiment, dtype: int64
```

Figure 5: Count of tweets corresponding to the positive and negative sentiments in dataset1.

```
In [12]: # Dataset 2
         data2.Sentiment.value_counts()
         # Count of tweets corresponding to the positive and negative sentiments

Out[12]: 1    790177
         0    788435
         Name: Sentiment, dtype: int64
```

Figure 6: Count of tweets corresponding to the positive and negative sentiments in dataset2.

Both datasets are balanced between the positive and negative sentiments of tweets. And that perfect to have a

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

clear and successful classification process. Be careful to always have a balanced dataset, otherwise the results will not be accurate and might migrate from positive and negative depending on the new data to test later, and that is something you do not want to happen in your projects.

Now we will check if there are any duplicate tweets. It is important to check for that because they might arise from Retweets by twitter users. And that should harm our data while training. And we will check the total number of Retweets in both datasets.

```
In [13]: #It is important to check if we have duplicates in tweets which is something that arise very often because of the RT (Retweet),
         # Show duplicated tweets if exist in Dataset1
         len(data[data.duplicated('SentimentText')])

Out[13]: 0

In [14]: # Show duplicated tweets if exist in Dataset 2
         len(data2[data2.duplicated('SentimentText')])

Out[14]: 0
```

Figure 7: Number of duplicates in tweets in both datasets

```
In [15]: # Display the number of RT in the first dataset
         CountofRT = data['SentimentText'].str.contains('RT').value_counts()
         CountofRT

Out[15]: False    10000
         Name: SentimentText, dtype: int64

In [16]: # Display the number of RT in the second dataset
         CountofRT2 = data2['SentimentText'].str.contains('RT').value_counts()
         CountofRT2

Out[16]: False    1578599
         True          13
         Name: SentimentText, dtype: int64
```

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Figure 8: Number of Retweets in both datasets

As we notice we have 0 duplicates in both datasets, but we have only 13 retweets in dataset2 whereas 0 for dataset1. For dataset2, we are lucky because even though there are retweets, yet there are no duplicates, and that's a plus in terms of training our data later.

We will be presenting the terminologies used in the preprocessing techniques in tweets presented in twitter application:

- Hashtags (#): This can be any word preceded by a #, once the user presses on the hashtag, all tweets that are related to this hashtag thus containing it inside the text, will appear.

- @mention: A mention can be done to mention a username on twitter application, for example mentioning President Joe Biden can be done like that: @JoeBiden.

- RT also called retweet: the procedure where a twitter user forwards or shares a tweet to another twitter user, or the newsfeeds is called Retweet. These retweets are used in spreading news and several important and interesting information on twitter, there will always be a main user that owns the post.
- Emoticons: often expressions that express certain mood of the twitter user and online scial media user in general for example: :O stands for surprised, :) stands for happy.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

We would need to have some data resources and dictionaries to help us during the data cleaning and preprocessing phase.

3.1.1- Data Dictionaries and Resources

- we have the **emoticon resource dictionary that** contains 132 most popular and known emoticons used online, with their corresponding positive sentiment (1) or negative sentiment (0).

Out[19]:

	Smiley	Sentiment
0	:-)	1
1	:)	1
2	:D	1
3	:o)	1
4	:]	1

Figure 9: Emoticon Dictionary

- The **acronyms data dictionary** that has 5465 data acronyms with their corresponding meaning and translation.

Out[20]:

	Acronym	Translation
5459	tomoz	tomorrow
5460	gpytfaht	gladly pay you tuesday for a hamburger today
5461	l8rz	later
5462	sase	self addressed stamped envelope
5463	bwoc	big woman on campus

Figure 10: Acronyms Dictionary

- The **stop words data dictionary resource** that contains all stop words which are considered unbeneficial in the text classification procedure and need to be removed.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Out[21]:

	Word
0	able
1	about
2	above
3	abroad
4	according

Figure 11: Stop words Dictionary.

- The **positive words and negative words resource dictionaries** with their corresponding sentiment positive or negative, these are considered highly useful in our work.

Out[23]:

	Word	Sentiment
2000	youthful	1
2001	zeal	1
2002	zenith	1
2003	zest	1
2004	zippy	1

Figure 12: Positive Words Dictionary

Out[24]:

	Word	Sentiment
0	2-faces	0
1	abnormal	0
2	abolish	0
3	abominable	0
4	abominably	0

Figure 13: Negative Words Dictionary

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

- A **Negations resource dictionary** that will help us indicate the negations in every tweet.

Out[25]:

	Negation	Tag
0	not	‖not‖
1	don't	‖not‖
2	doesn't	‖not‖
3	aren't	‖not‖
4	isn't	‖not‖

Figure 14: Negations Dictionary

The resources we have will help us remove some of the complexities in the data, a lot of negations, acronyms, positive and negative words are used in the dataset, knowing that the negative and positive dictionaries will boost the F1 scores and accuracies. The data resource dictionary that corresponds to stop words, has over 635 words. we will see about that later, during the testing phase of our models before choosing the best one for testing it on new data.

5.2- Sentiment Analysis of Social Media Pictures and images:

In previous research, machine learning was used to predict the sentiment of a picture. Certain datasets were there for training and testing. At start I used the dataset and code presented in **[90]** Results will appear later in this report. But on the other hand, after doing more research, I was able to figure out the presence of the deep learning library which is DeepFace **[16] [17] [18].** In this case we do not need any data, its already built in inside this library, all we need to do, is use this libraries tools, and input any picture, and it will provide us with the sentiment analysis of this picture along with other criteria as well. Comparison of both approaches will be delivered (Between Machine learning and Deep Learning)

5.3- Sentiment Analysis of Meeting or Interview videos:

No datasets are needed, the only data we need is a video (mp4 file used here), and with the help of FFmpeg

and moviepy libraries **[48] [49] [50] [51]** we can proceed with the preparation of our video to start detecting sentiments. We will start by importing from moviepy.video.io.ffmpeg_tools the tool which is ffmpeg_extract_subclip in my case, I made a sentiment analysis of Joe Biden's 60-minute 2020 Election interview **[91],** so I chose my time intervals accordingly, at each question I took a sample. In my case I chose the two parts where Joe biden was asked "Do you think trump would win?" and the second question was "Do you think raising taxes is a good idea in this crisis? "So, I chose my time intervals accordingly (This could differ depending on the target of the user and what time slot he/she wants to have full detection precisely on) the function ffmpeg_extract_subclip takes 4 parameters, the original video, the start time, the end time, and the destination of the first sample of the video. It will generate a new video under these time slots. The starting point and ending point take a value n as seconds, so if you input 5 in the starting time and 10 in the end time, it means that you will receive a new video taken from the original video that was between 5seconds and 10seconds. If the start time is at minute 5, then the user should input 5*60 which is 300seconds.

First, we start by choosing our video, in my case its Joe Biden's Interview video.

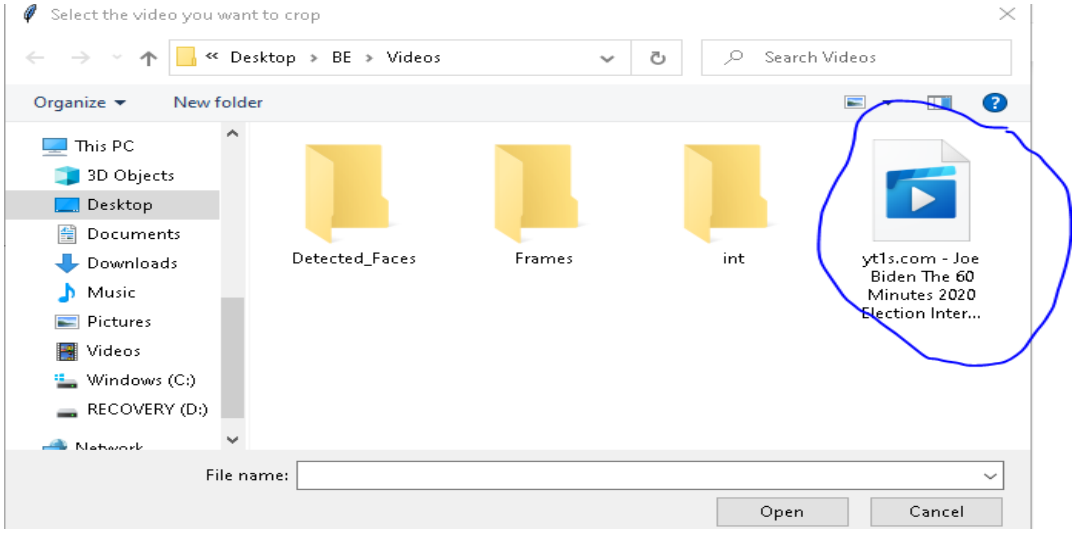


Figure 15: Choose the main video.

After that, the system will ask you for the number of videos you want to create, in our case we have 2, since we will be analyzing two questions being asked to Joe Biden.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

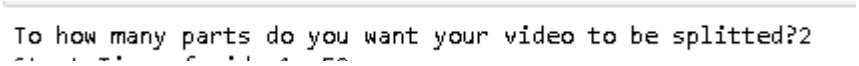


Figure 16: Identifying the number of videos to split.

Then the system will ask the user to input the parameters of each video, in our case we have two videos, the system will ask the user to input the parameters for each one, we will be showing the first one. In Fig18, we can notice that the start time is 58seconds and the end time is 65seconds which is one minute and 5 seconds (Certain procedure will be applied in the future work to do that automatically without inputting the timing). These are the start and end time of the first question we wanted to analyze in this interview. Then the destination of the video with its name is brought to the user, for example I saved the video is "BidenVideo1" and do not forget to write .mp4 so it would be saved as an mp4 format file.

Figure 17:Parameters of the first video

Same procedure is applied to the remaining videos, in our case there is only one video left so its only once again to perform the same procedure and input the parameters. And then we will have the two cropped videos ready for later testing and sentiment analysis.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

5.4- Sentiment Analysis of vital signs:

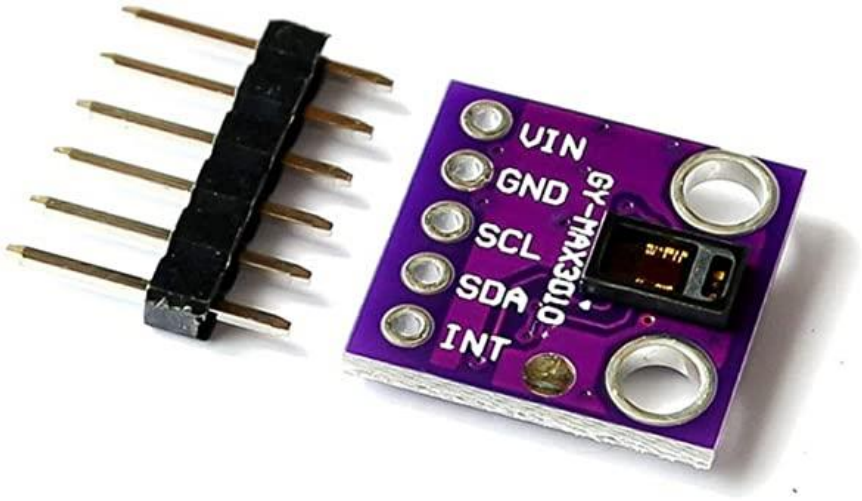


Figure 18: Sensor Chip

This is the Max30102 sensor that I got from Katranji EKT**[98]**. A chip and male headers. First, I connected the male headers through the holes of the chip to the breadboard. The sensor did not work so I had to head back to the store. I went back and I soldered the openings between the male headers and the holes so it can be 100% connected. After the soldering is complete, I went back to connect the sensor.

Oxygen Saturation Pulse Oximeter measures oxygen saturation. Before we understand the working principle of a pulse oximeter, we need to understand what oxygen saturation is. Oxygen enters the lungs and then into the blood. The blood brings oxygen to all organs of our body. The main route of oxygen delivery the blood flows through hemoglobin. We call oxygen-free hemoglobin oxyhemoglobin (deoxyhemoglobin). Oxygenated hemoglobin, we call it oxygenated hemoglobin (Oxy Hb).

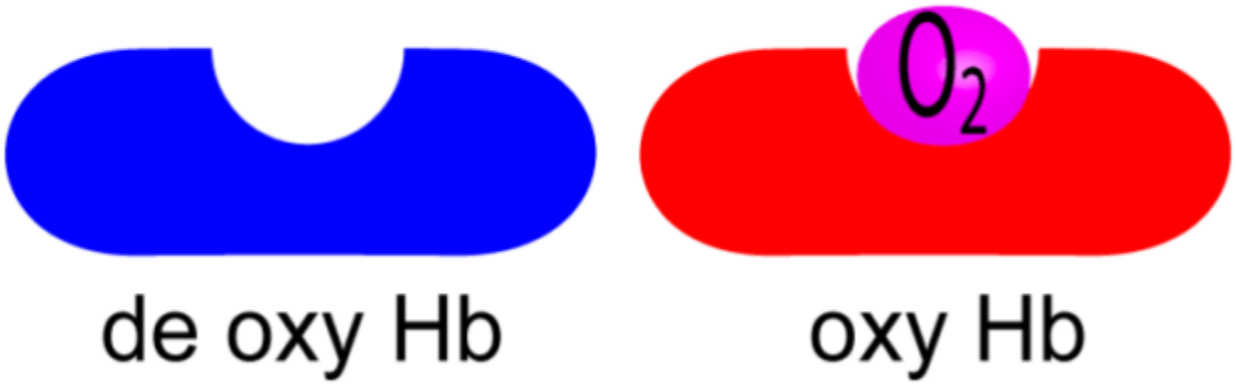


Figure 19:oxy-deoxy hemoglobin

Oxygen saturation simply represents the percentage of available hemoglobin that carries oxygen. Please take the following situation**.** There are 16 units of hemoglobin, and none of the 16 contains oxygen. Therefore, the oxygen saturation is 0%.

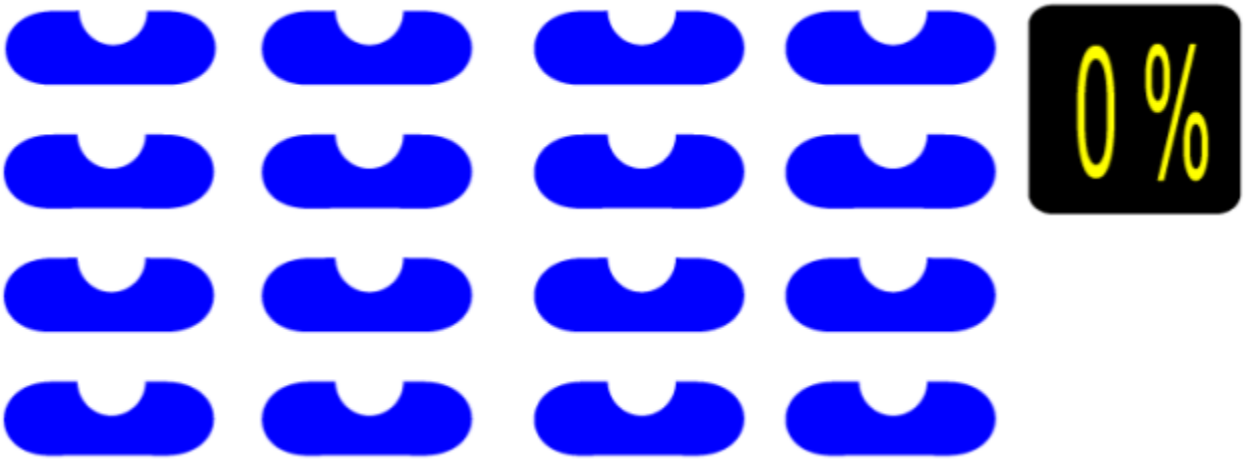


Figure 20: No oxygen carried.

And if it carries the 75% oxygen the saturation will be 75%.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

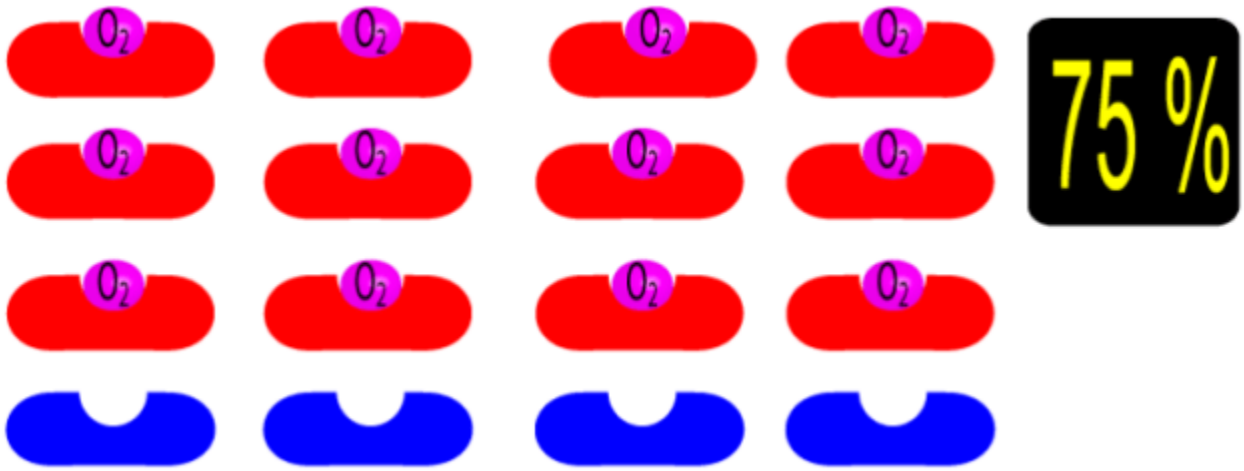


Figure 21: 75% oxygen carried.

And if all the blood cells carry the oxygen, then of course it is going to be 100%.

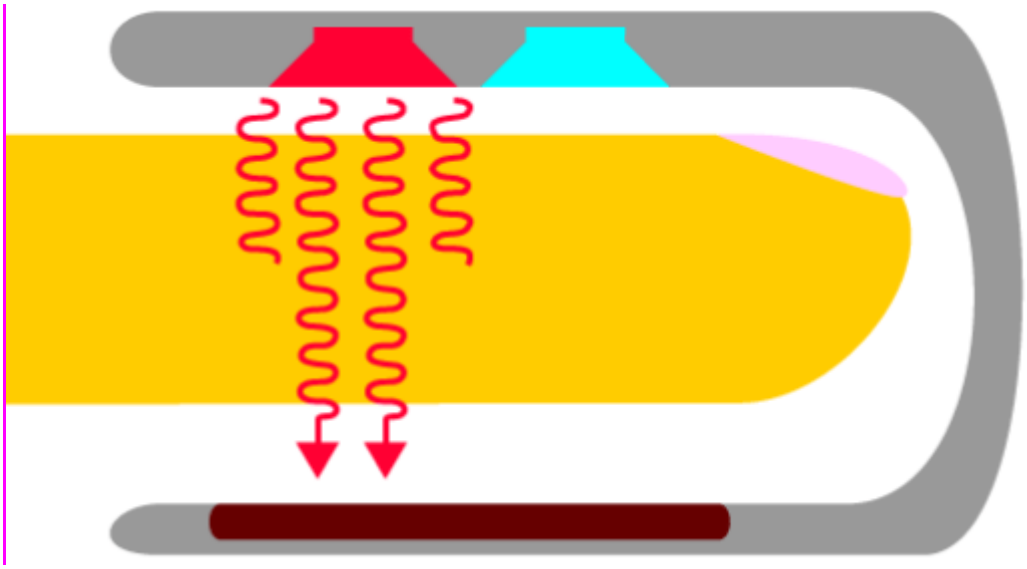


Figure 22: Oximeter methodology

There are many shapes and sizes of oximeters, but the general principles of pulse oximeters remain the same. Some of them are like the fixture type shown in the image above, while others are using

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

reflection techniques like the ones I use. Work in reflection. Therefore, when light is transmitted or reflected, we can measure the oxygen concentration.

Maxim MAX30102 includes a high-sensitivity heart rate monitor, pulse oximeter and MAX30102 accelerometer. MAX30102 is an integrated heart rate biosensor and pulse oximetry measurement module. The MAX30102 ambient light suppression electronics is a complete system solution that simplifies the design process of mobile and portable devices. The main function works with extremely low power consumption. Possibility of fast data output. Resistant to motion artifacts**. Additional features.** Decided. According to the I2C standard**,** an independent 5.0 **V** power supply for the internal LED Tiny 5 interface.14-pin optical module 6 mm x 3.3 mm x 1.55 mm integrated **s**afety glass for best and reliable performance**.** Very low power consumption operation for mobile devices**.** Programmable sampling rate and LED power for energy-saving low heart rate monitors Output **(**<1 mW> ultra-low-low trip current (usually 0.7µA), fast data output, high sampling rate, reliable resistance to motion artifacts, high signal-to-noise ratio. The module can be turned off by software**.** Zero standby current makes the bus always on. -40°C to +85°C operating temperature range **[109]**

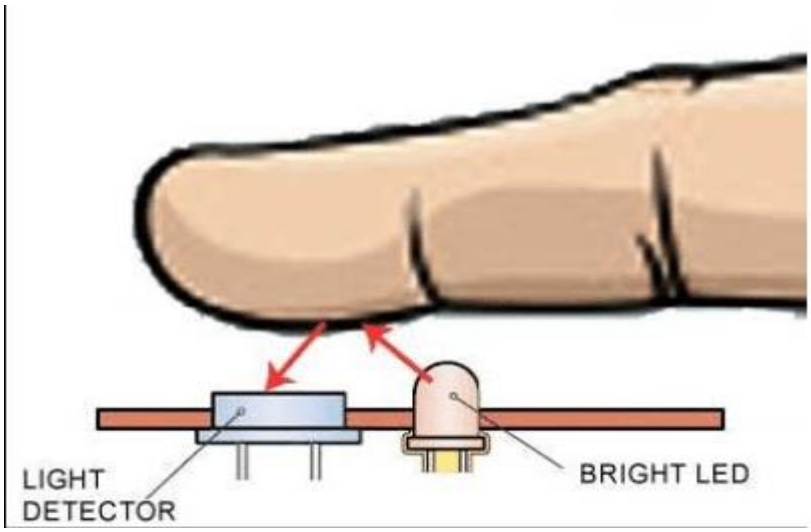


Figure 23: Heart rate recognition

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Heart rate sensor: When your finger is placed on the heart rate sensor it sends a digital heatstroke signal. When the heart rate sensor is working, the heart rate indicator light will flash at the same time for each heartbeat. The output can be directly connected to the microcontroller to measure the number of heart beats per minute (BPM). Its principle is to use each pulse to modulate the light in the blood flowing through the finger. The sensor consists of an ultra-bright red LED and a light sensor. The LED must be very bright because the maximum light must pass through the finger and be detected by the detector. When the heart pumps the pulse through the blood vessels, the finger becomes opaquer, so less light reaches it. The signal from the detector changes with each heartbeat. This change becomes an electrical pulse. The signal is amplified and controlled by an amplifier, which outputs a +5 V logic level signal. The output is also indicated by an LED, which flashes on every heartbeat.

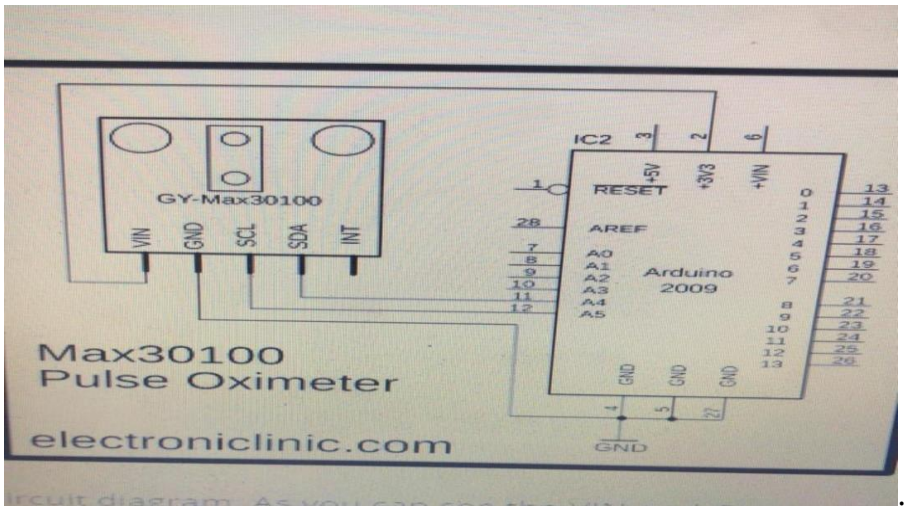


Figure 24: Connections

This is how we connected the sensor to the Arduino Uno.

Vin ⟶ 5V
GND ⟶ GND
SCL ⟶ A5

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

SDA ⟶ A4
INT ⟶ Not connected

The first code used was from the library given by (Kontakt/Max30100). The code was verified and uploaded successfully. But there was a problem in initializing the sensor. On the serial monitor the output was as given in the below picture.

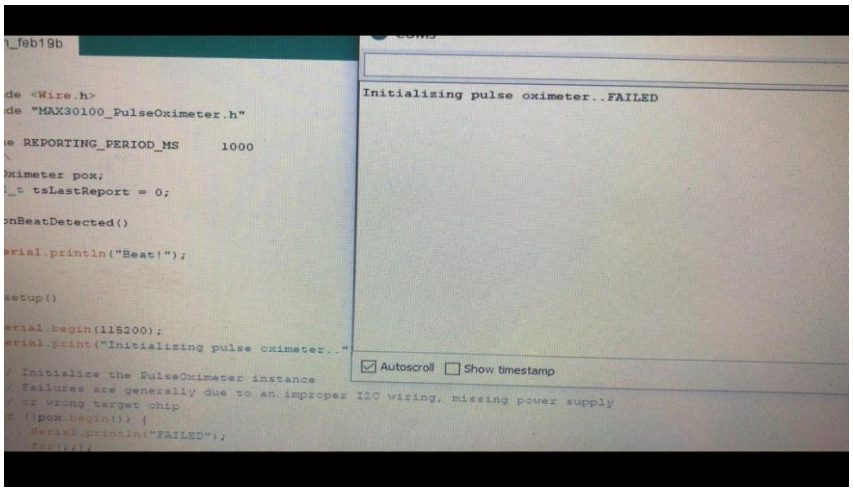


Figure 25: Initialization failed.

A problem in initializing the sensor was found and thus no results for the heart rate and the spO2 were found. After this problem emerged, detection of the sensor address needed to be made. So, I went to File ---> example ---> wire ---> i2c scanner and ran the code of this example. As seen in the picture below, the device was found on the 0*57 address.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

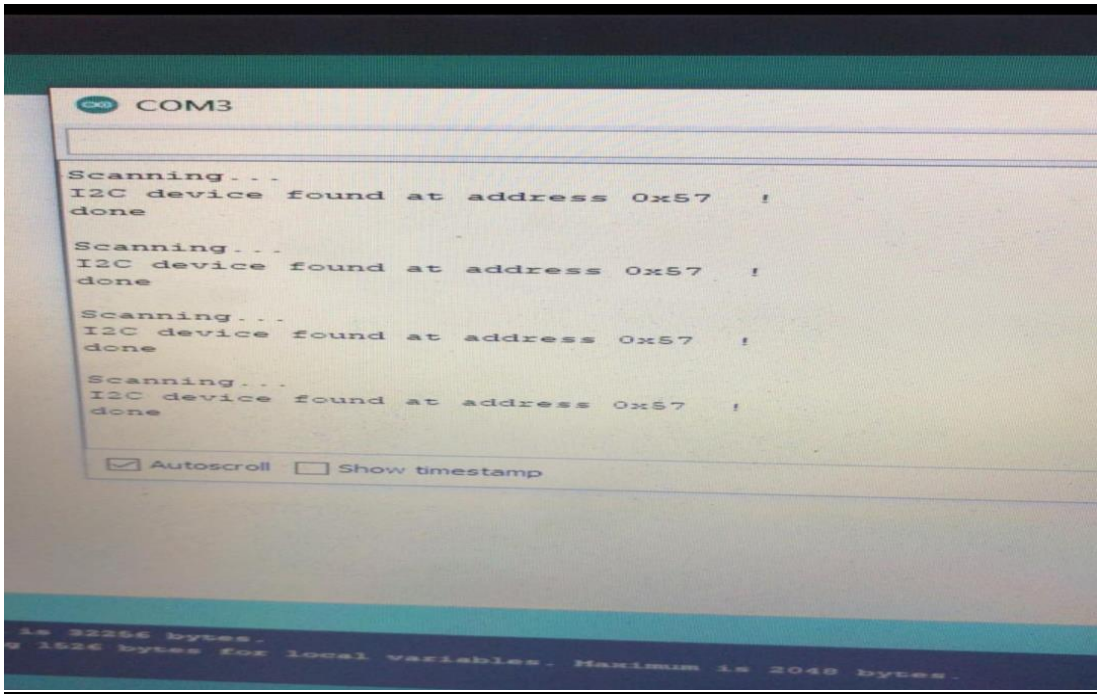


Figure 26: Device found on address.

After checking the availability of the sensor, we knew that the problem was not from the sensor itself and that the issue was from the address. So, to change the address, a new software was introduced “VScode”. This software will let us roam freely in the library without the restrictions of the arduino IDE software.

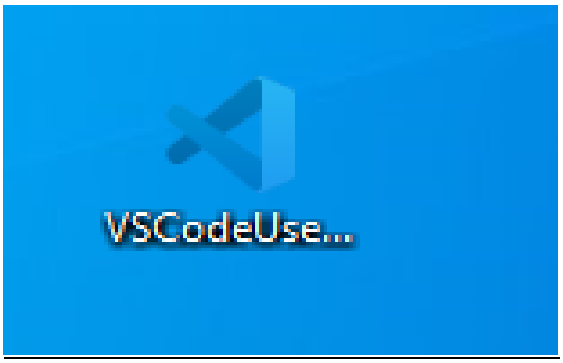


Figure 27: VScode app

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

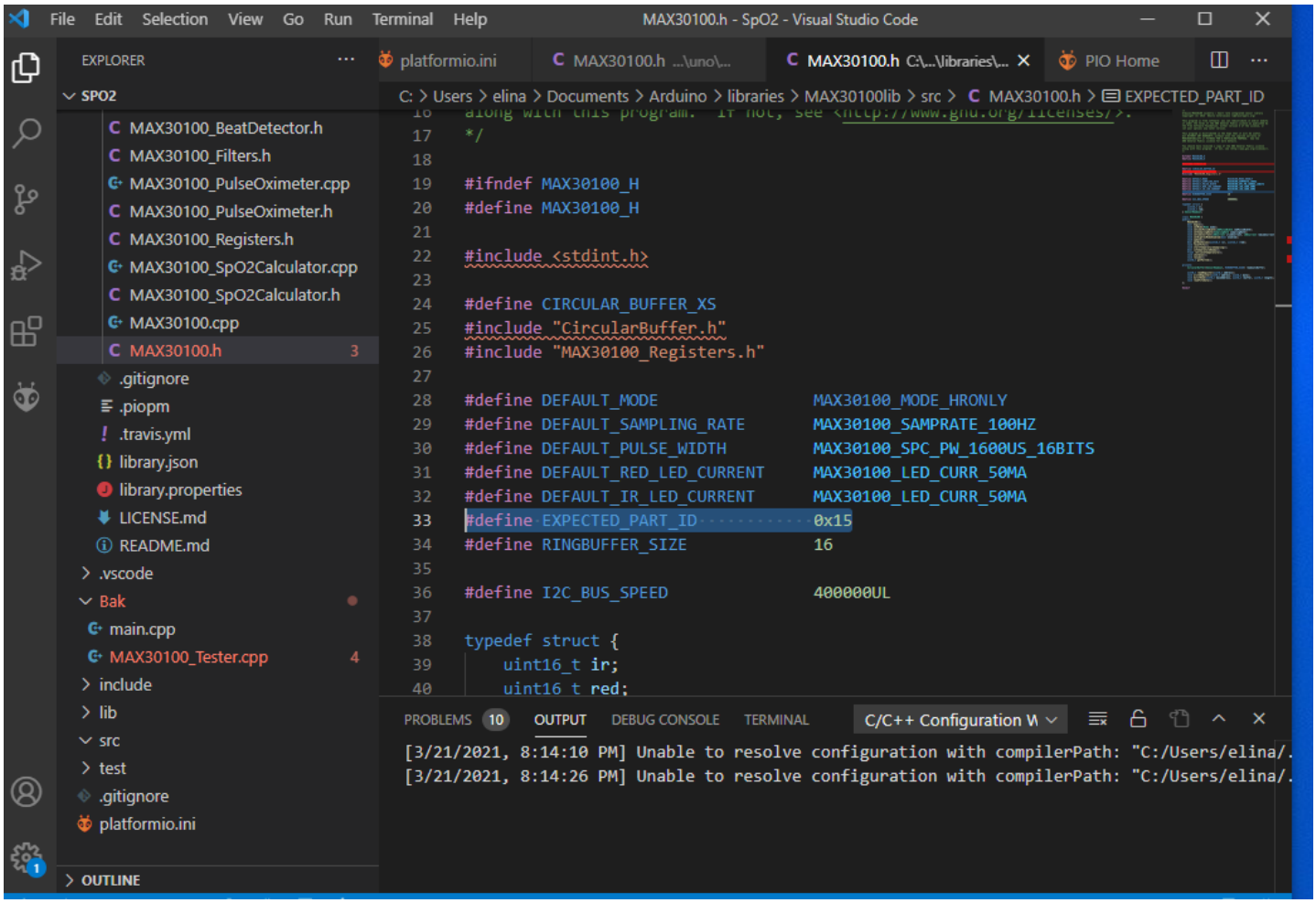


Figure 28:Changing address

Visual Studio Code is a free source code editor developed by Microsoft for Windows, Linux and macOS. Features include debugging support**,** syntax highlighting, smart code completion, code snippets**,** code refactoring and embedded Git. This software has an extra edge over the arduinoIDE by the free roaming in the libraries. Without this software,

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

changing of the address of the sensor would not have been possible. **[95]**

In Fig.27, the library was uploaded and as seen on the left of the picture, the Max30100 .h file was opened. In this code we changed the address of the sensor to 0*15 which is the expected value. This value was obtained from the .cpp file.

After changing the address, we then inserted this code into the Arduino software and connected the system; initialization of the sensor was success this time.

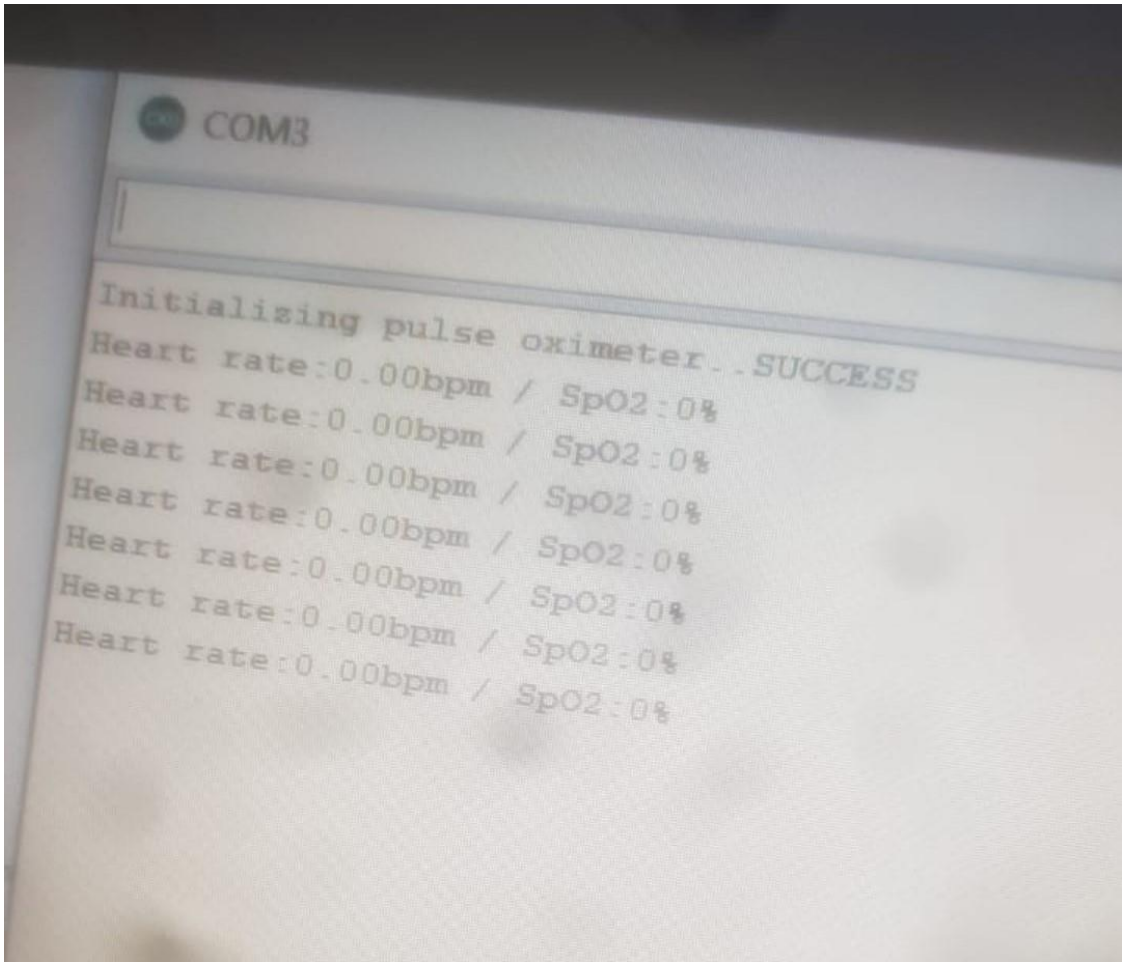


Figure 29: Initializing success.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

It is true the sensor was successfully initialized but, no results were found. The heart rate and the SpO2 were reported as zeros all the time. And that is when we knew that there should be a change in the whole library. After trying a couple of libraries, we found one from SparkFun_Max3010x. So basically, this library works for the different versions like Max30100, max30102 and max30105. After uploading this library from sparkfun, the sensor turned on and results started to be recorded. **[101]**

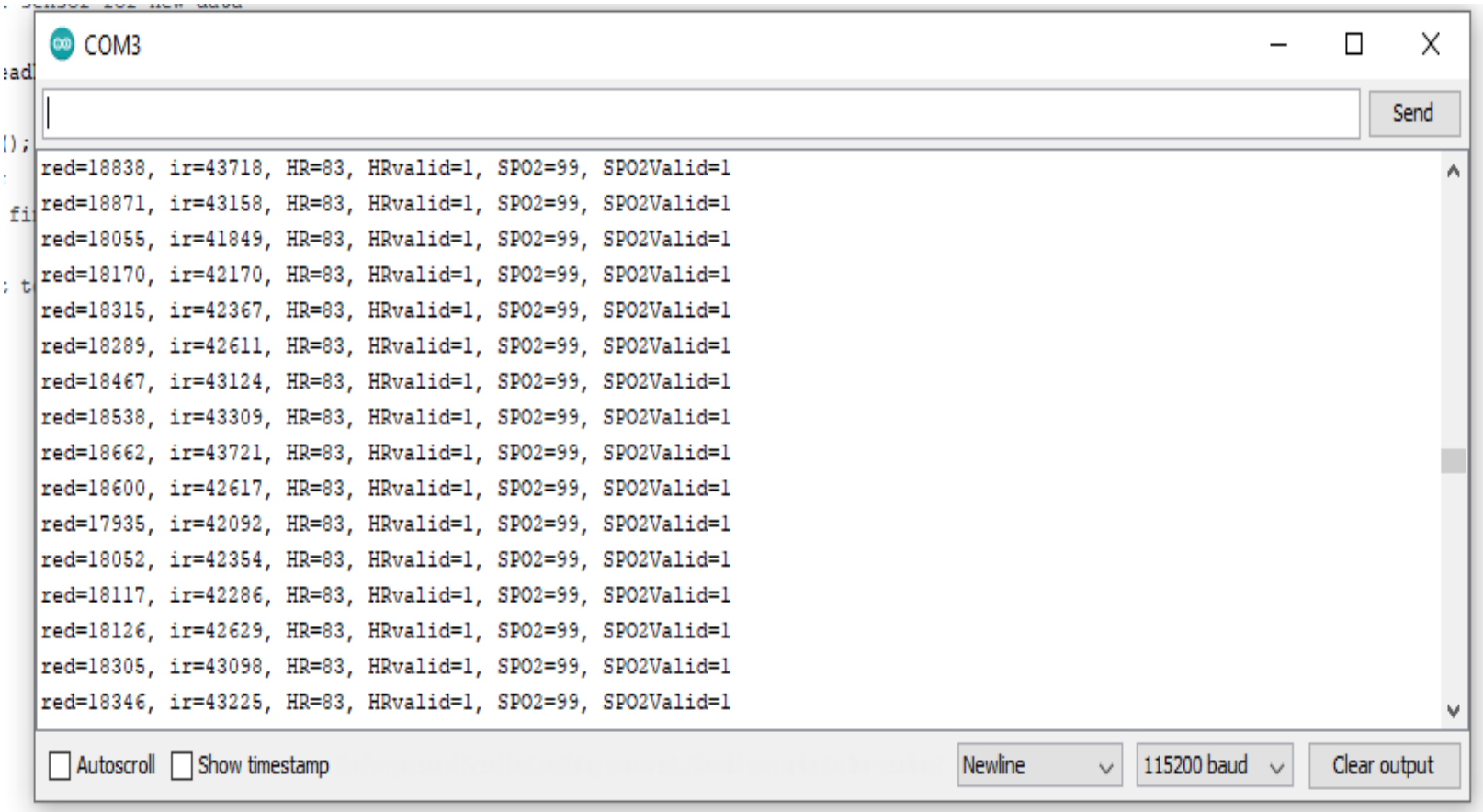


Figure 30: Sensor started recording.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

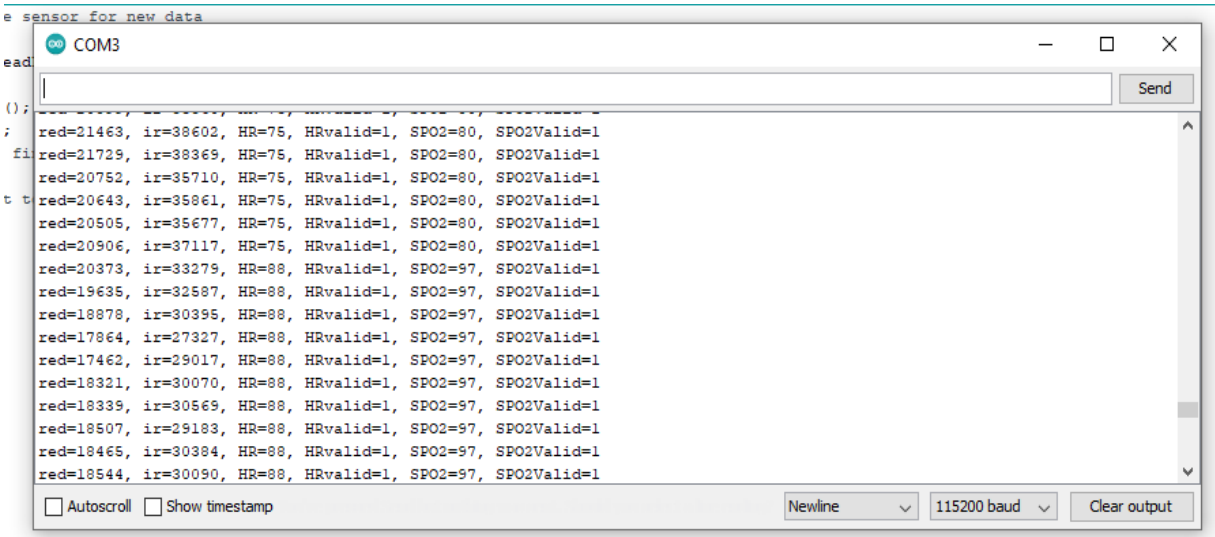


Figure 31: Another snip of recordings

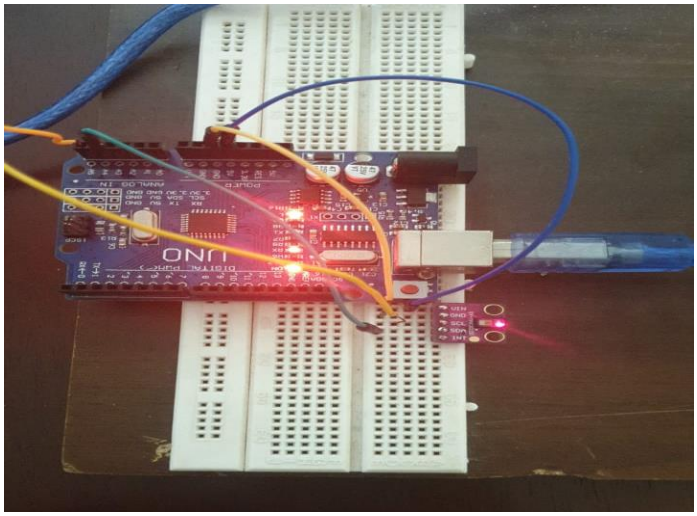


Figure 32: Sensor without finger

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

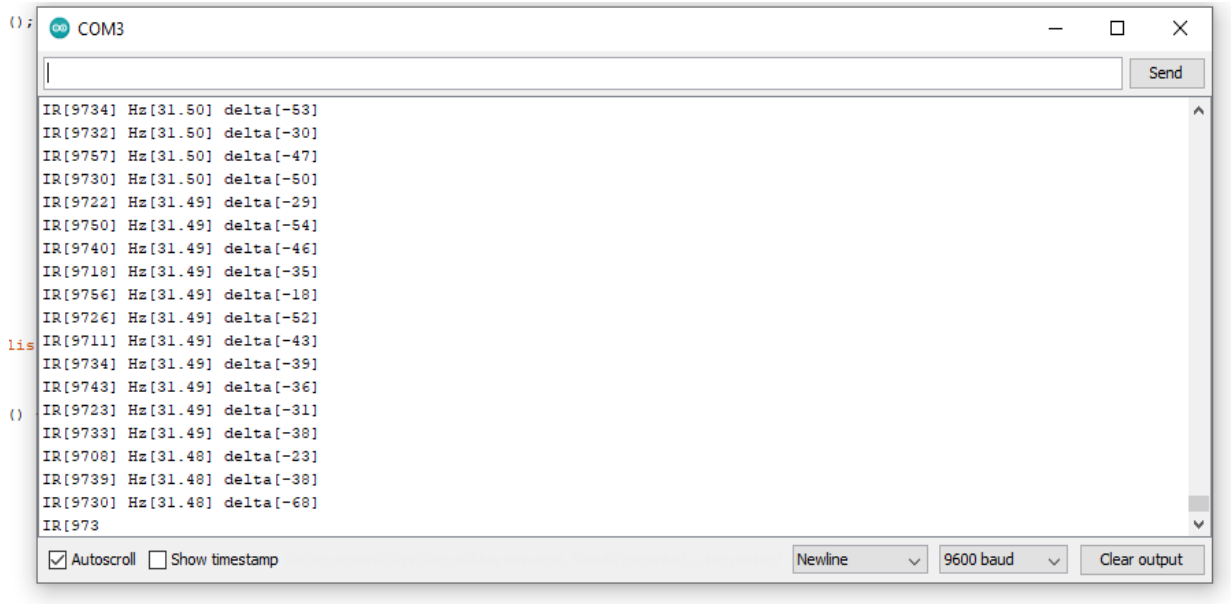


Figure 33: No finger detected

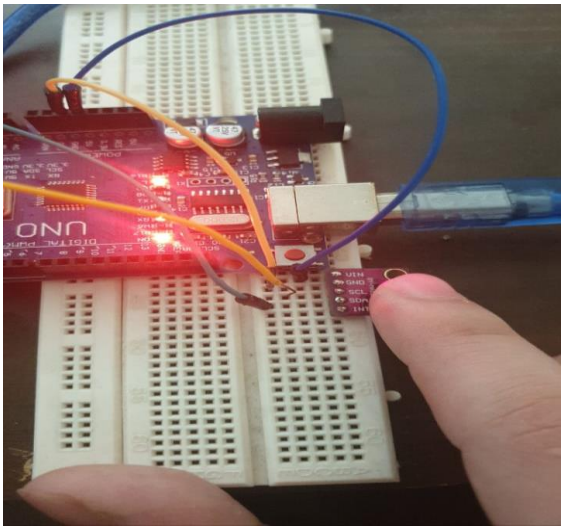


Figure 34: Sensor with finger

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

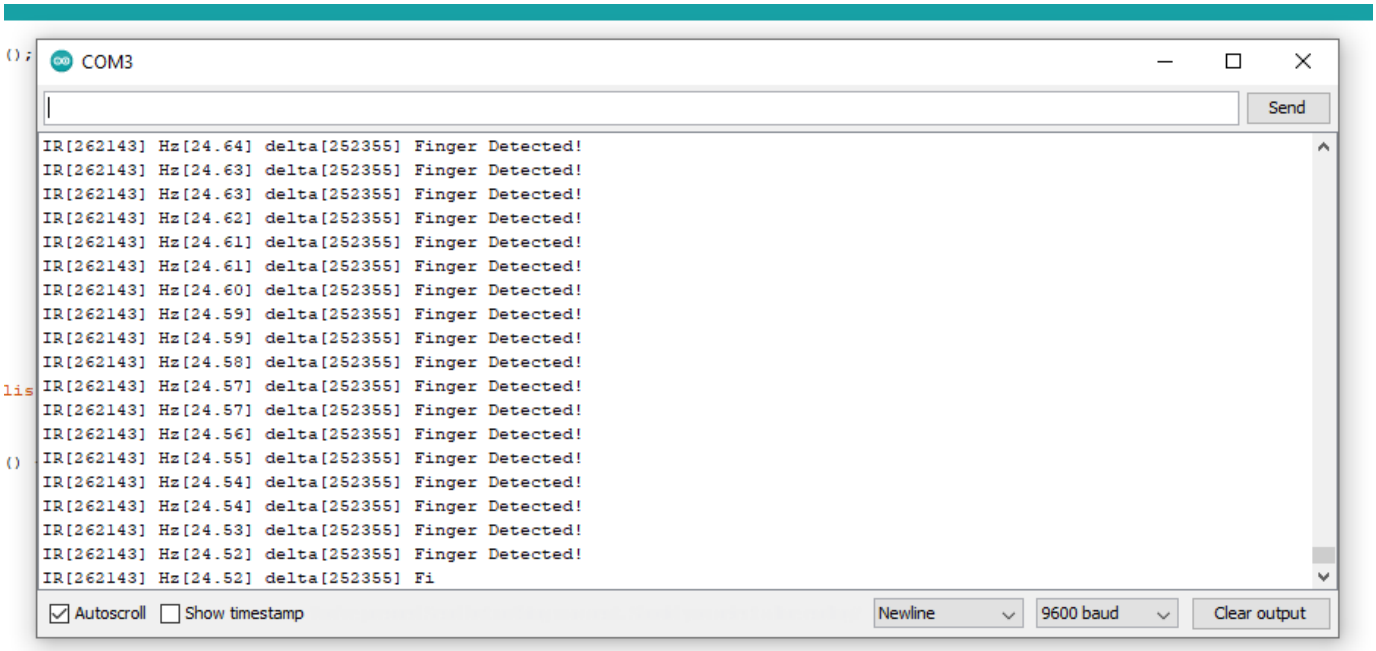


Figure 35: Finger detected.

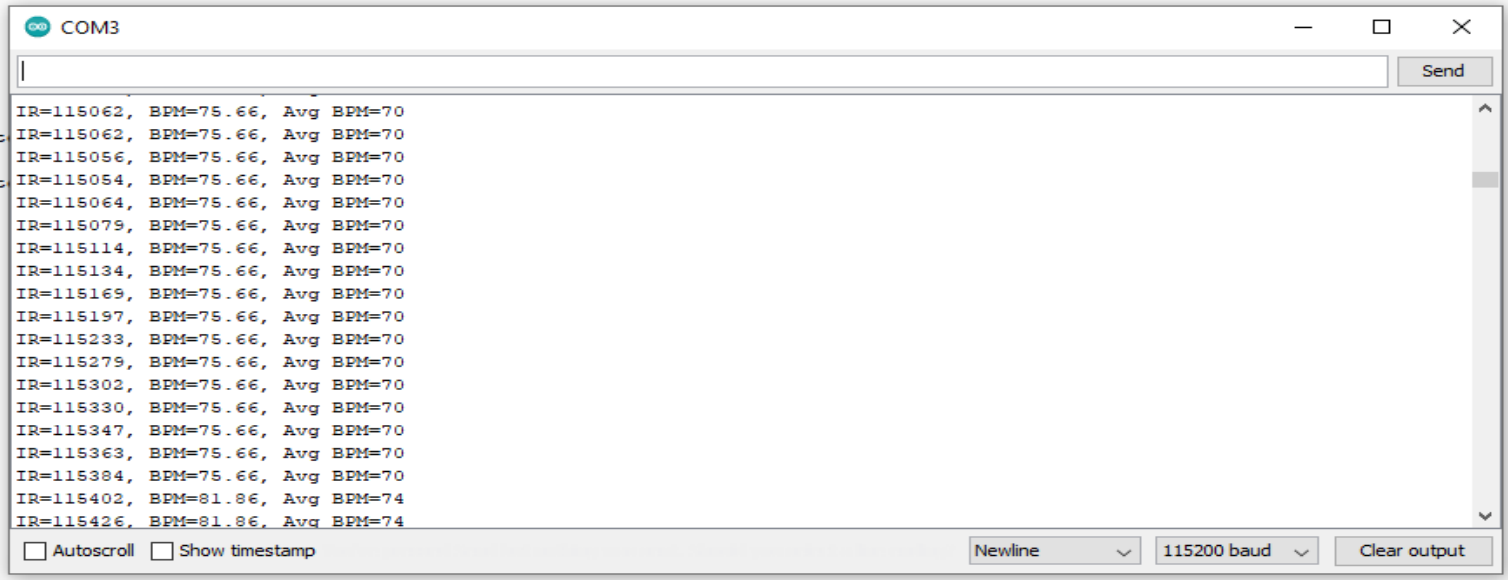


Figure 36:Heart rate recordings

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

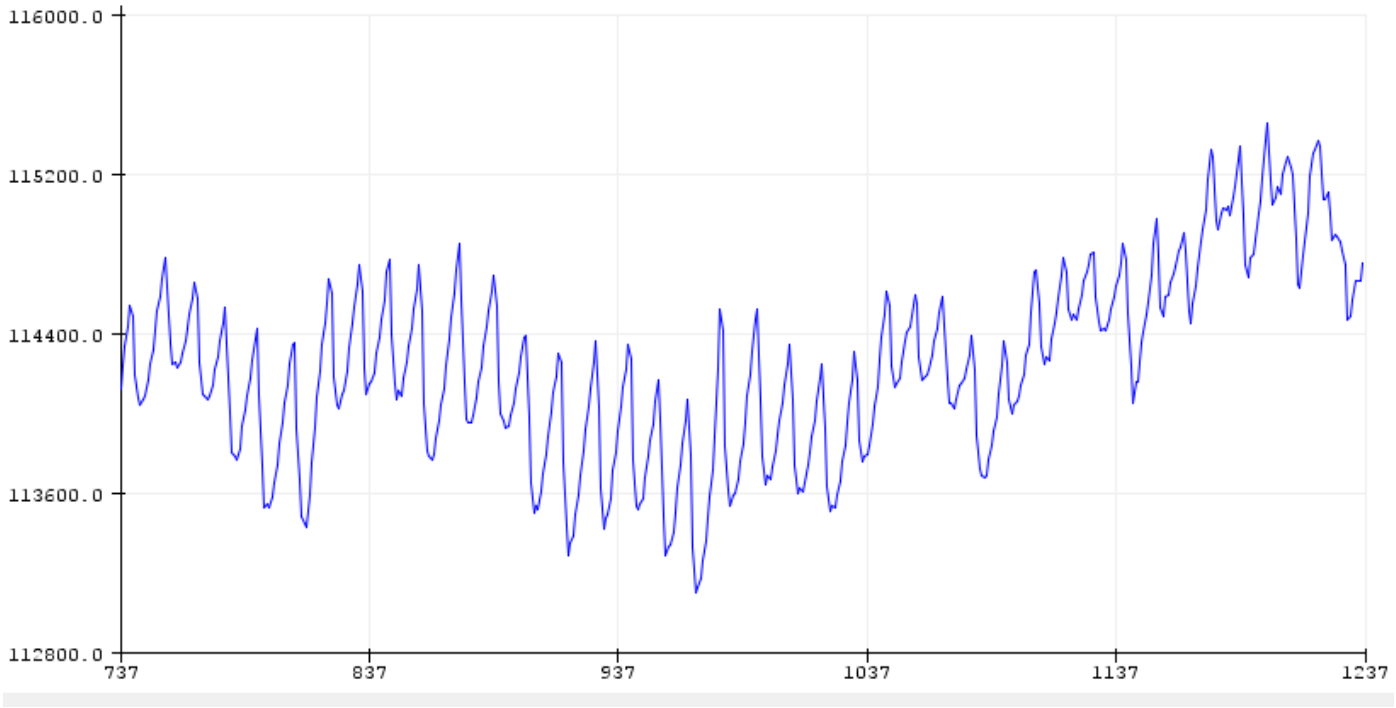


Figure 37: Beat plot.

Dataset:

After managing to make the sensor work. 10 subjects were contacted to take their beat plot for further analysis.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Subjects	10
Age	18 ~ 25
Avg. Weight	78 kg.
Avg. Height	175
Healthy	10

Figure 38:Table showing Subject's description.

Video:	Sentiment:	Description:	Duration:
Video 1	Neutral	Delivery man doing his job	1 min.
Video 2	Angry	Southpaw: Aggressive boxing scene	1 min.
Video 3	Sad	Southpaw: Daug. Heart Break	1 min
Video 4	Happy	Newborn laughing	1 min.

Figure 39: Table showing video description.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

General Steps:

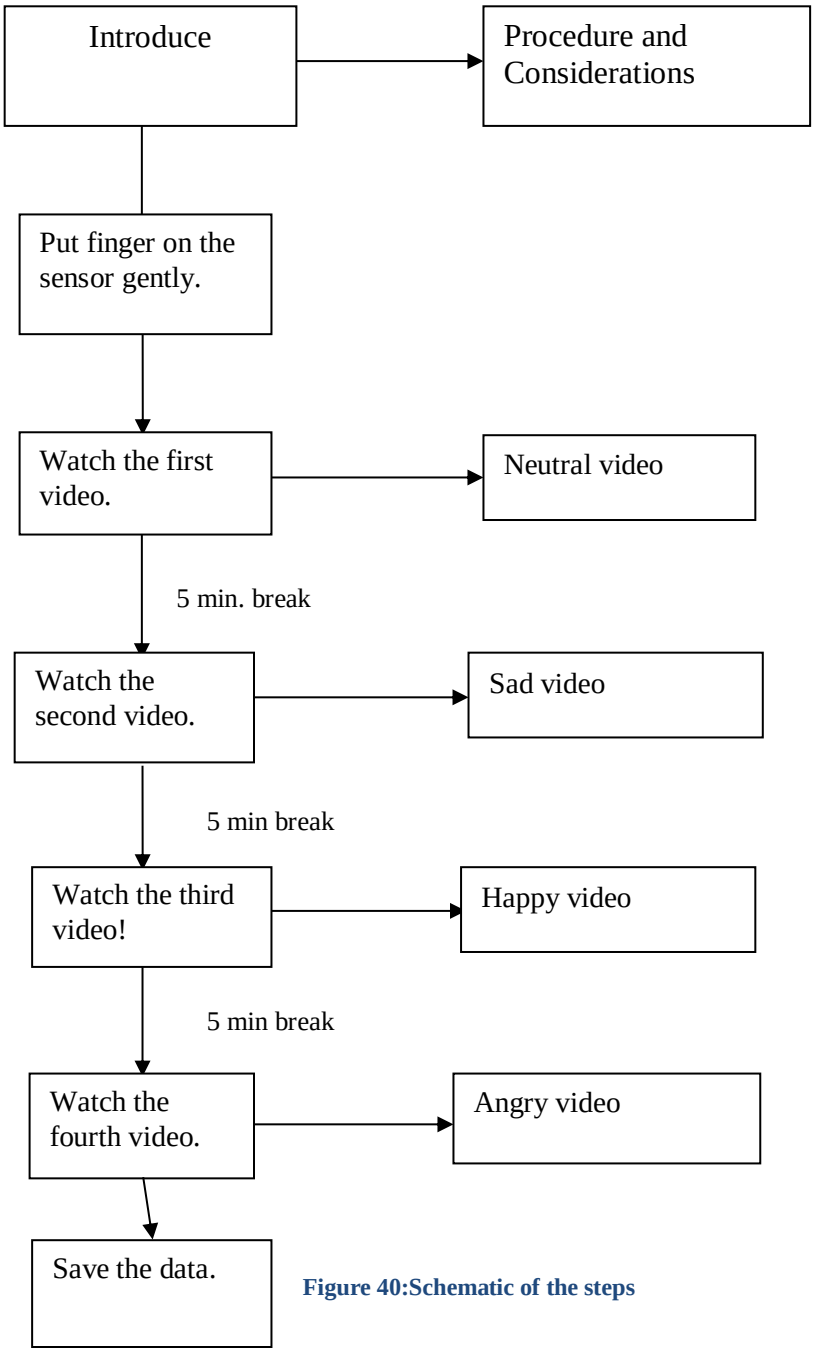


Figure 40:Schematic of the steps

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Subjects were asked to watch the videos and put their finger on the sensor to monitor their heart activity. After monitoring, results for the beats/minute were recorded as the following.

Sub/Vid :	Video :	1(Neutral)	2 (Sad)	3 (Happy)	4 (Angry)
Subject :					
1		63	58	76	85
2		70	65	78	74
3		57	60	66	71
4		81	77	78	90
5		68	66	64	70
6		55	63	72	80
7		65	59	63	74
8		76	71	80	92
9		64	60	71	81
10		86	84	90	102

Figure 41: Table showing subjects with sentiments.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

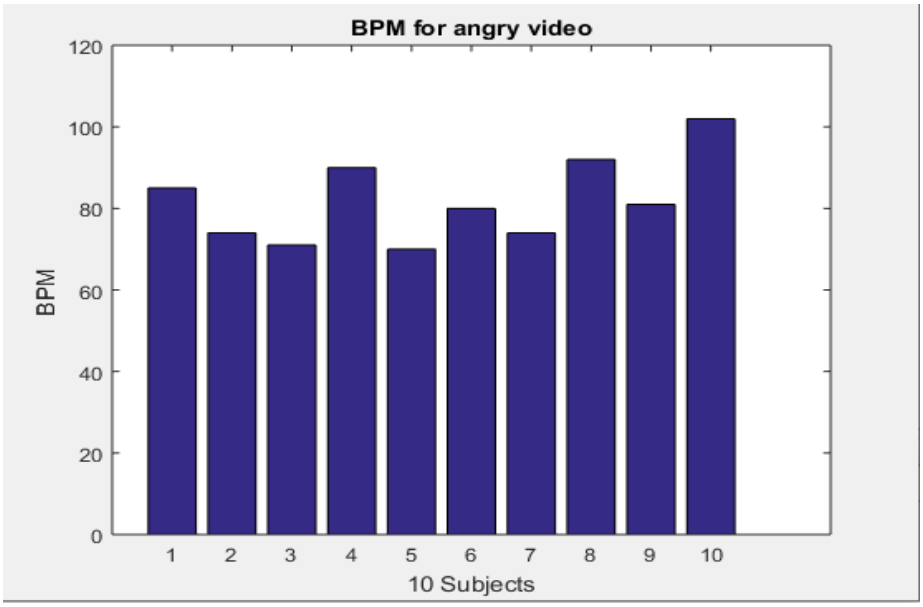


Figure 42:BPM for angry video subject

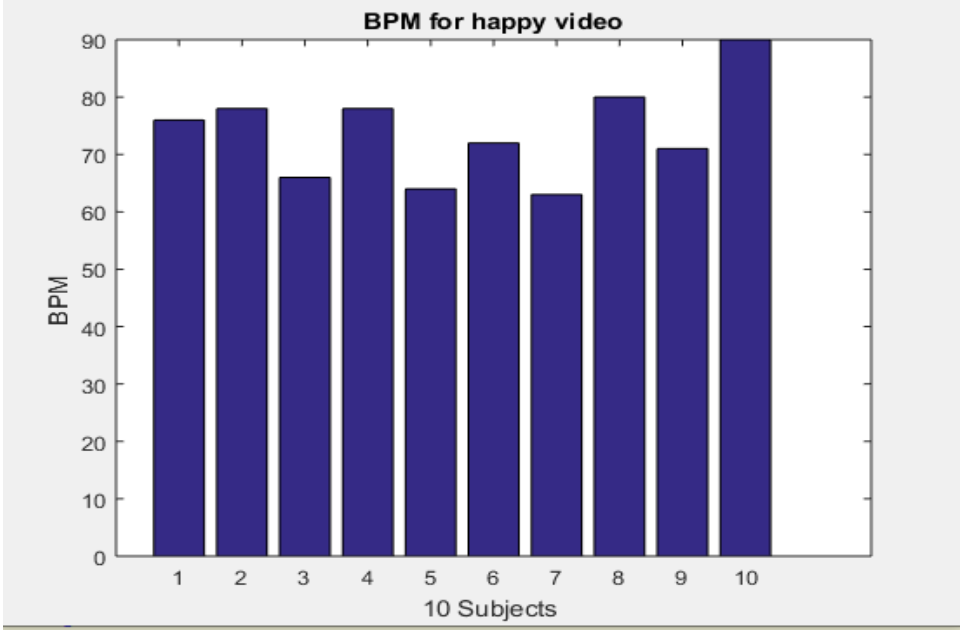


Figure 43: BPM for happy video subject

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

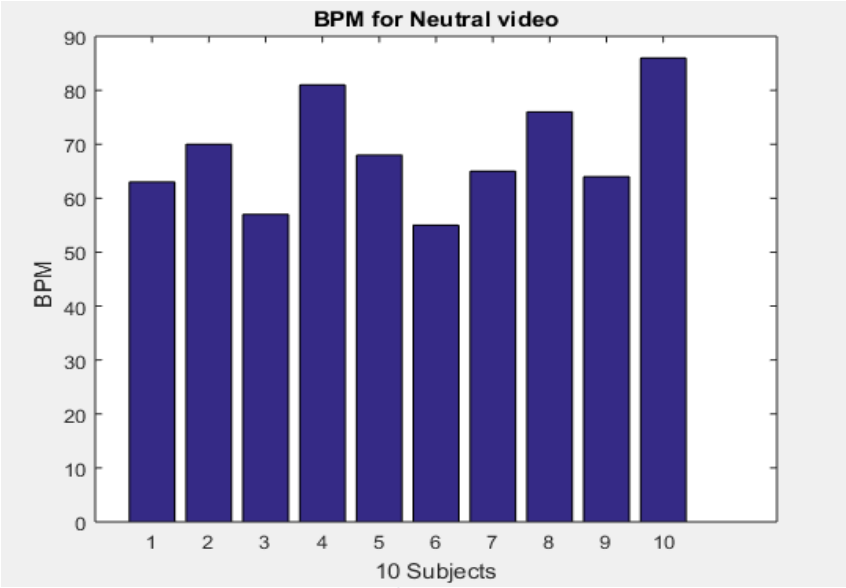


Figure 44: BPM for neutral video subject

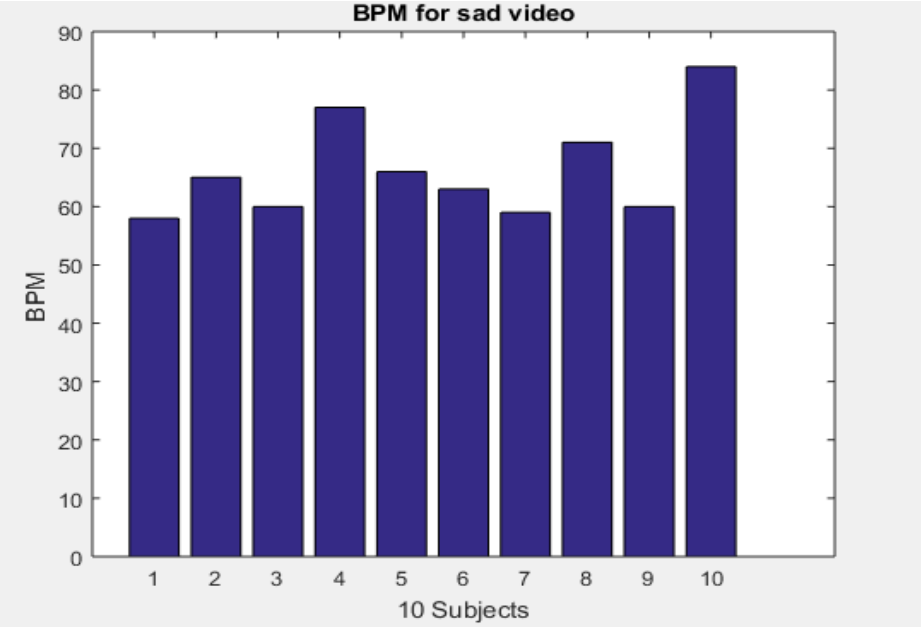


Figure 45: BPM for sad video subject

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

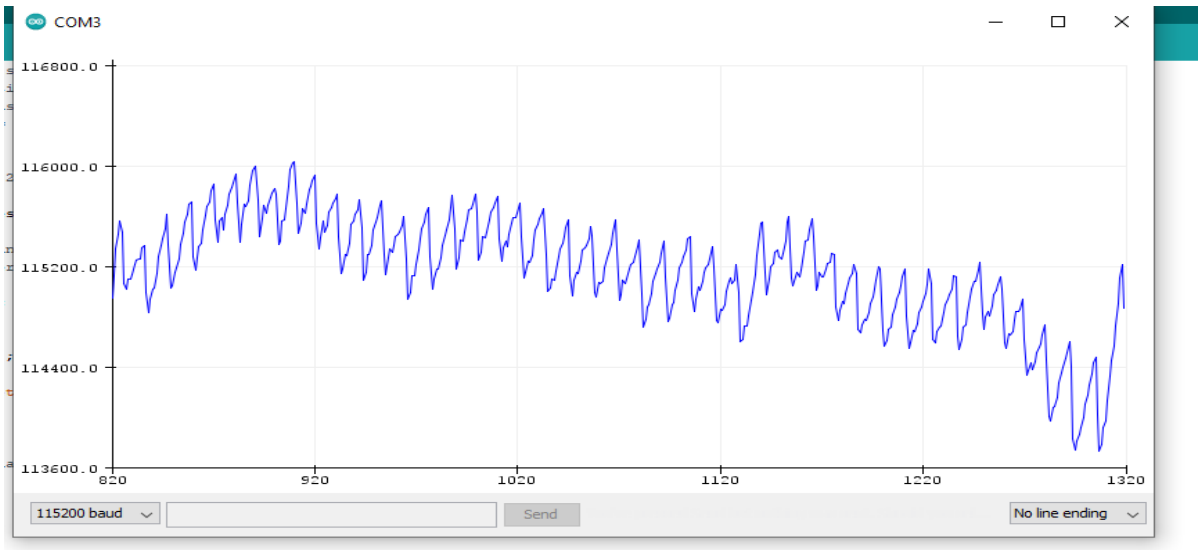


Figure 46: Sad plot

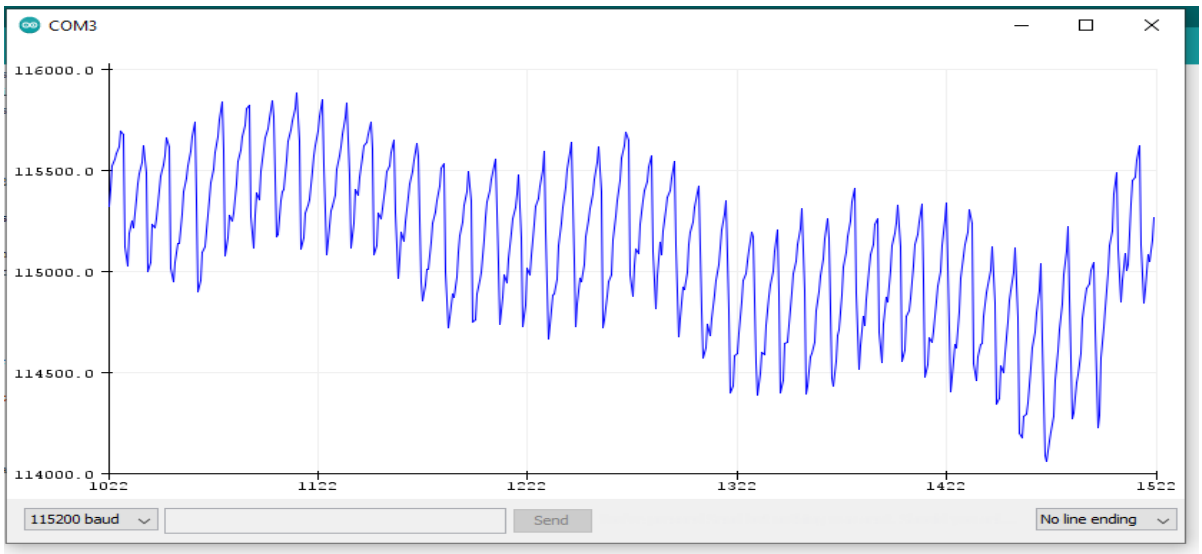


Figure 47: Neutral plotting

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

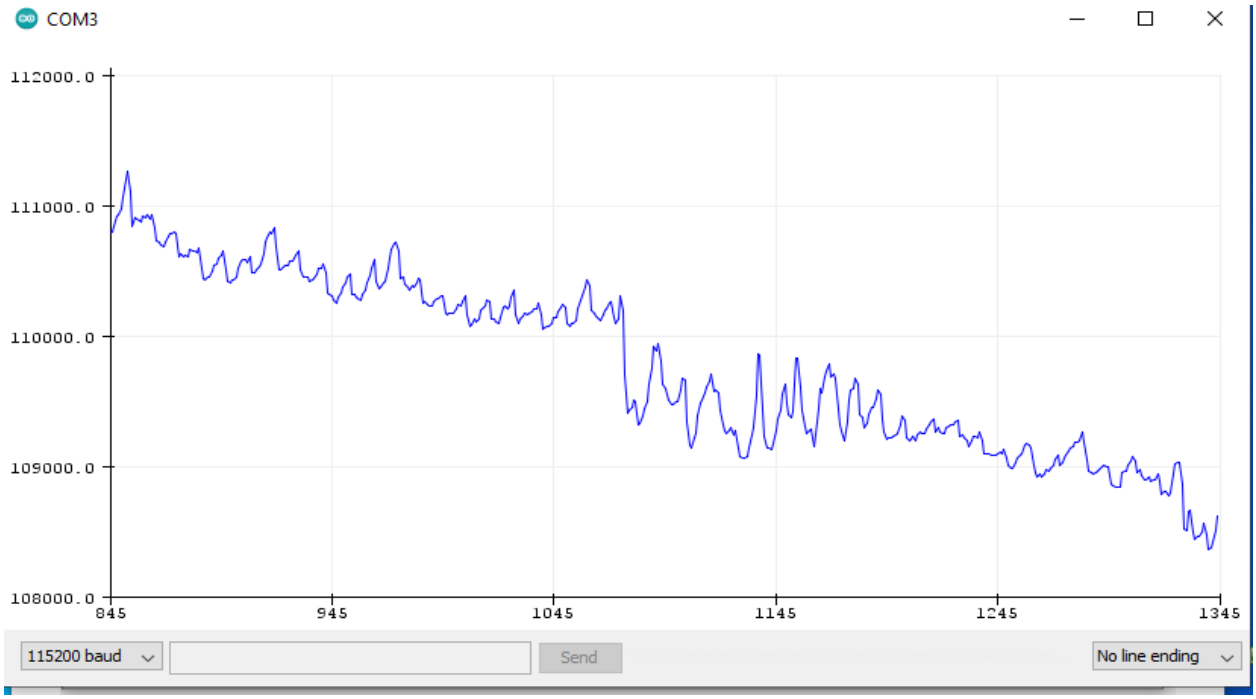


Figure 48: Happy plotting

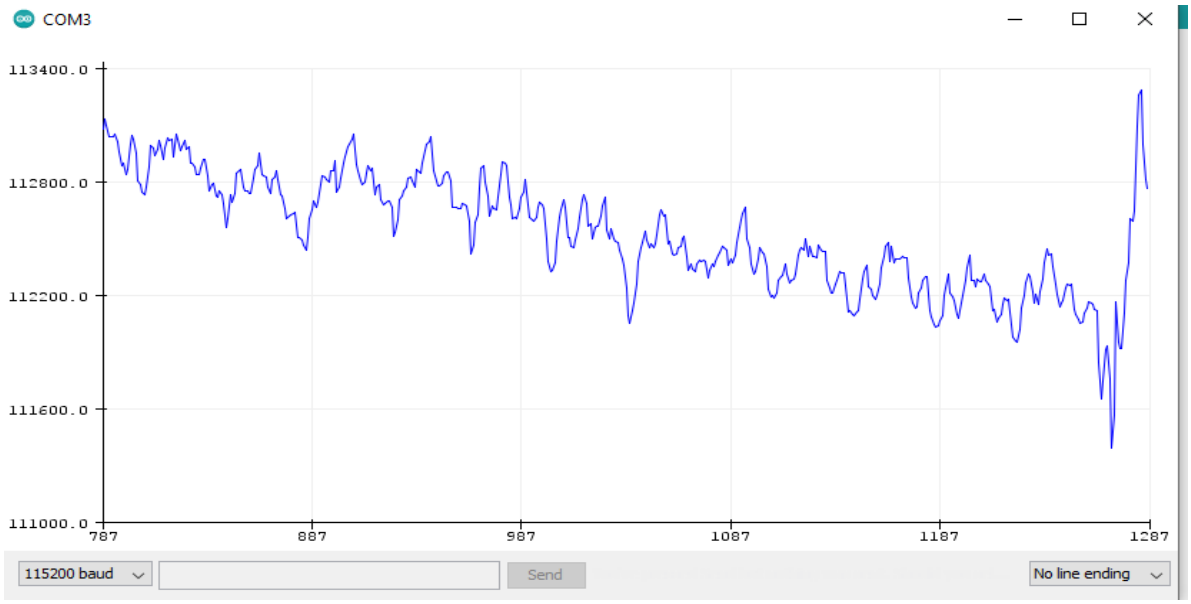


Figure 49: Angry plotting

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

After the collection of the signals, IR values were collected and saved in a .txt file for each subject for each video. This leaves us with 10 (subjects) × 4 (videos) = 40 .txt file filled with the IR values.

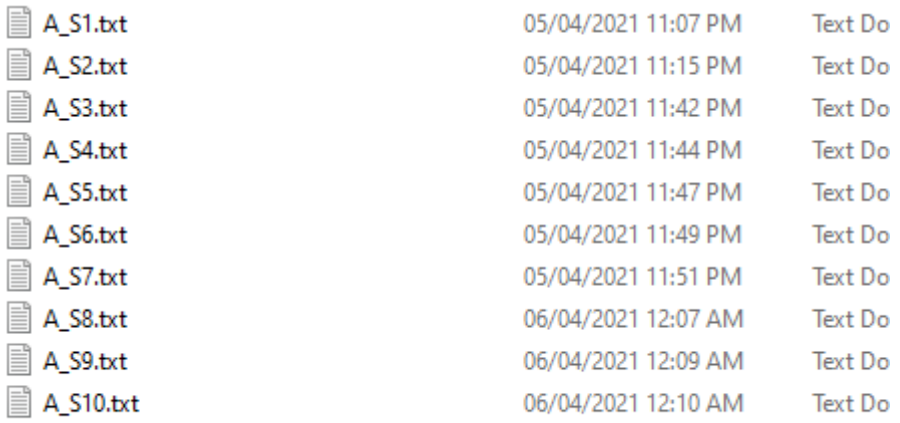
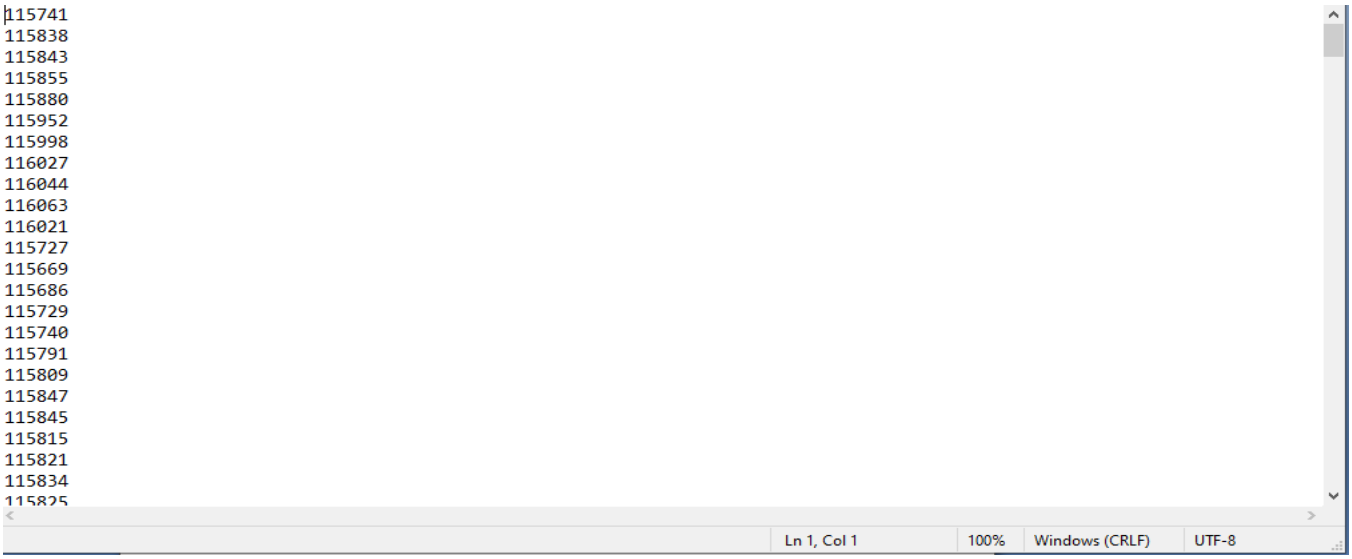


Figure 50: IR values for subjected who watched angry video.



Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Figure 51:IR values for subject 1 for the angry video.

After the collection of the IR values, features should be collected from them. So, an excel sheet was made to organize all the information.
Features:

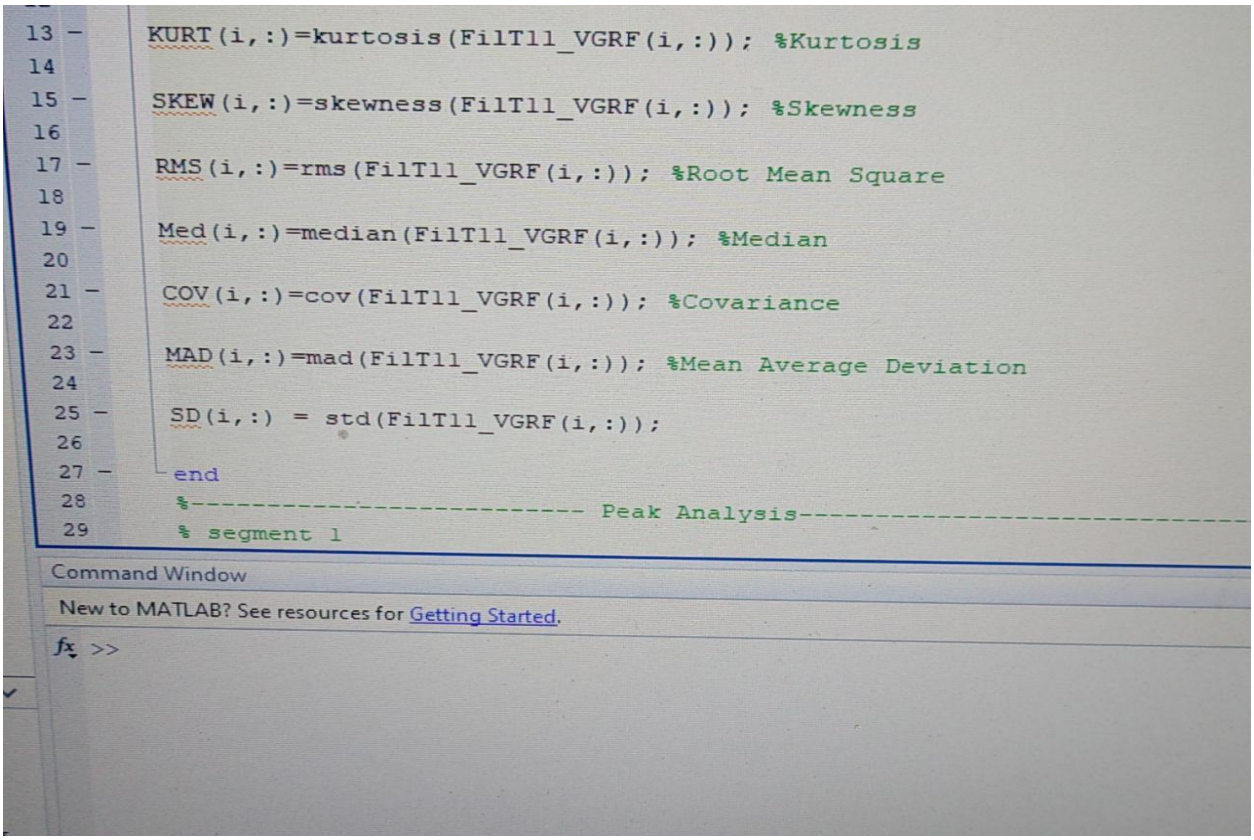


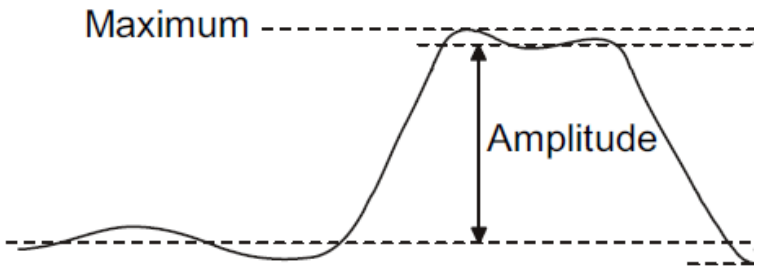
Figure 52: Matlab code to calculate the features.

Features that were calculated are: Maximum, minimum, median, kurtosis, skewness, root mean square, Mean Average Deviation, Covariance, Standard deviation. BPM was also added next to each subject as it adds extra significance.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

1- Maximum: The maximum feature signifies the highest point in the signal. **[110]**



2- Minimum: The minimum specifies the lowest point in our signal. **[111]**

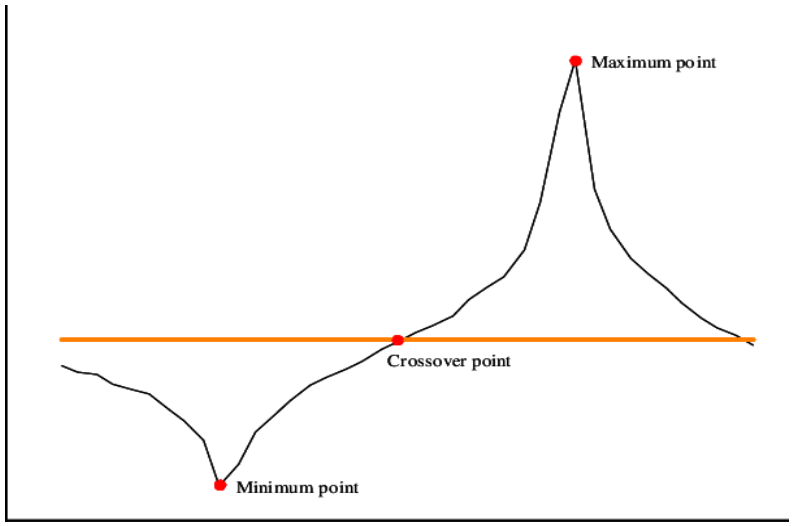


Figure 53: Minimum of a signal

3- Median: The median is the value when half of the observations in the sample are below the median and half of the observations are above the median. The median is often used as a measure of the "center" of the signal because it is less sensitive to outliers. **[112]**

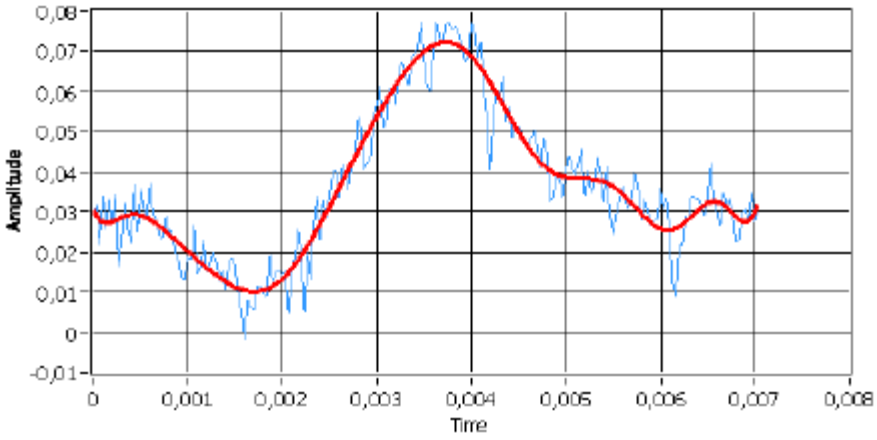
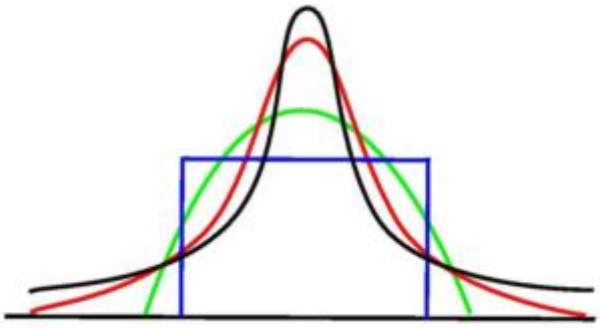


Figure 54: Median of a signal

4- Kurtosis: Kurtosis is a statistical parameter used to characterize a signal. Basically, it provides a measurement of the "peak" of a random signal. The higher kurtosis signal has more peaks than three sigma. In other words, the peak value is more than three times the root means square value of the signal. **[113]**



Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Figure 55: Kurtosis

5- Skewness: is a measure of symmetry in a distribution. Measure the probability number in the tail. This value is usually compared with the kurtosis (3) of the normal distribution. The tails of the data set are heavier than the normal distribution (more tails). **[114]**

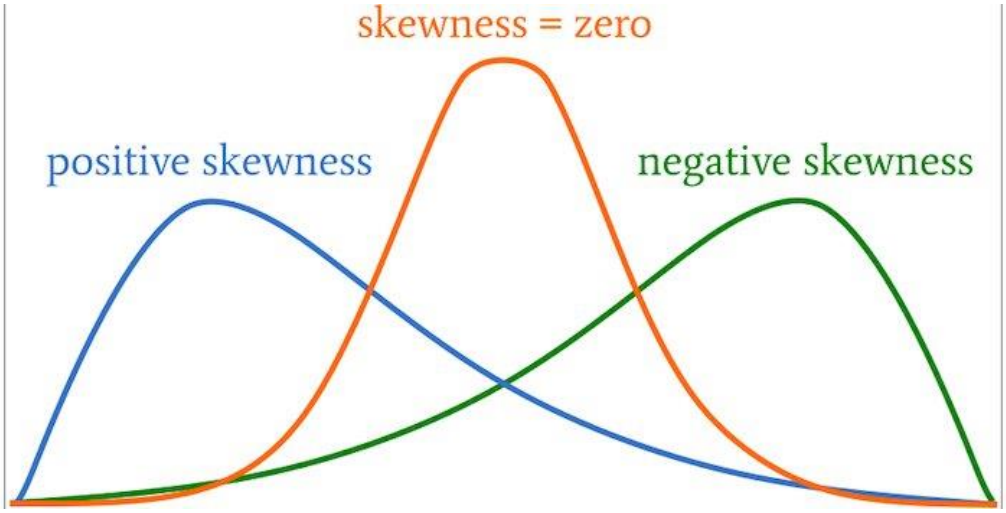


Figure 56: Skewness feature

6- Root Mean Square: RMS is a measure of the effective value of a signal over time: it is not an "average" voltage, and its mathematical relationship to the peak voltage depends on the type of waveform. **[115]**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

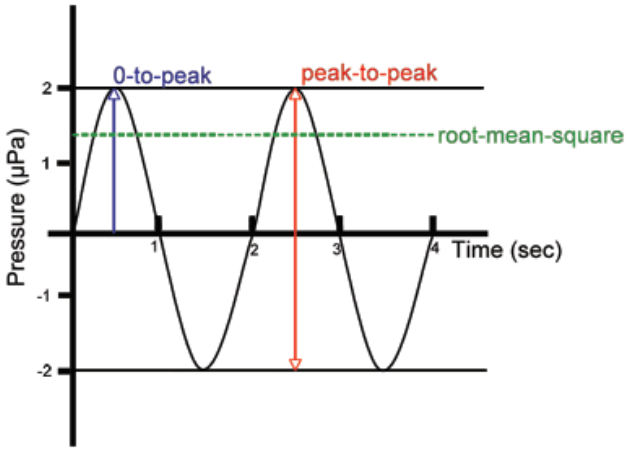


Figure 57: RMS of a signal

7- Mean Average Deviation (MAD): The average signal deviation is determined by adding the deviations of all individual sample values and dividing by the number of sample values N. Note that we take the absolute value of each deviation before summing. Otherwise**,** the positive and negative terms will average zero**.[116]**

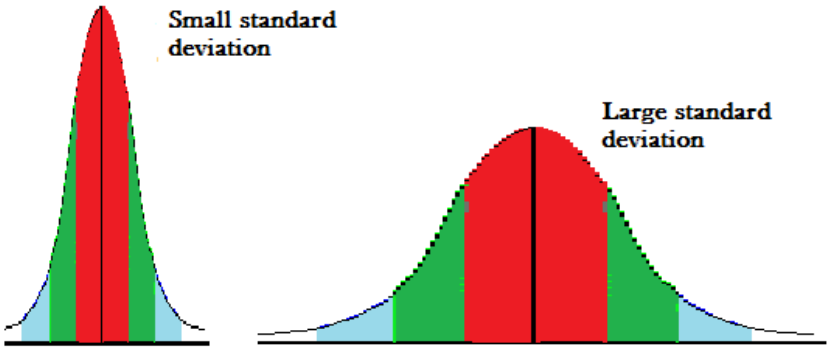


Figure 58: Mean average deviation.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

8- Covariance: Covariance measures the directional relationship between the returns of two assets. Positive covariance means that asset returns will move together, while negative covariance means they will move backwards. **[117]**

Figure 59: Covariance

9- Standard deviation: The standard deviation (or σ) is a measure of how scattered the data is relative to the average. A low standard deviation means that the data is grouped by means, while a high standard deviation means that the data is more scattered.**[118]**

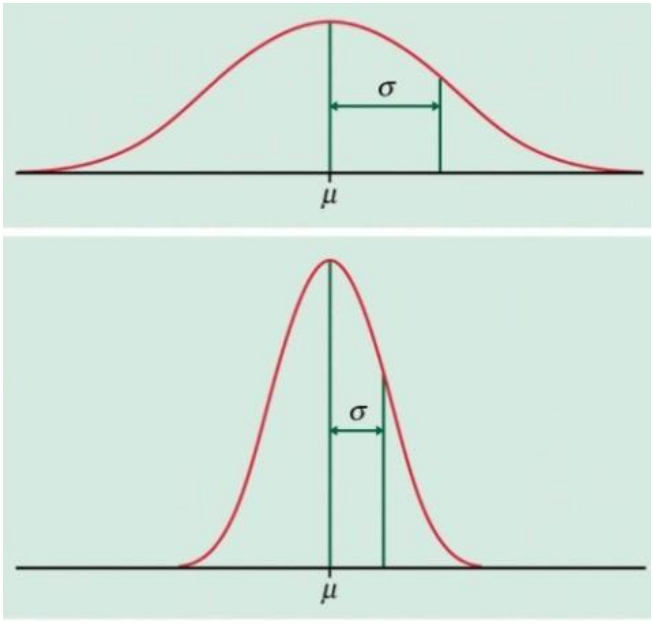


Figure 60: Standard deviation

10- BPM: And since we are working with a system that needs to classify the data, BPM was also added next to the features to add extra significance to each subject.

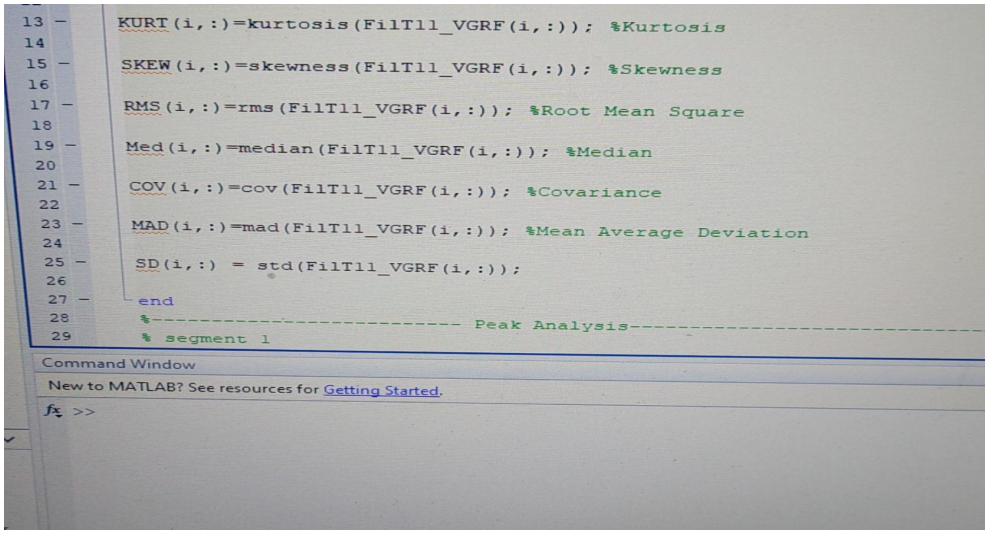


Figure 61: Matlab code to calculate the features.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

MATLAB was accessed to calculate the features. **In figure.61**, code is shown that is responsible to calculate certain features; kurtosis, skewness, rms, median, covariance, the mad and standard deviation. The IR values were then imported to the MATLAB workspace one at a time to calculate all these features from a certain (.txt) file. When all the files were imported and all the features were collected, every value was recorded on an excel sheet table.

	MAX	MIN	MEDIAN	KURT	SKEW	RMS	MAD	COV	SD	BPM	Response
A_S1	116063	113425	114744	2.0829	0.7087	1.14E+05	650.1017	5.78E+05	760.4468	85	Angry
A_S2	116801	114422	1.16E+05	2.1661	0.0955	1.16E+05	484.5664	3.43E+05	585.9303	74	Angry
A_S3	117951	111358	114654	2.55	-0.2614	1.12E+05	281.0022	1.25E+05	353.0927	71	Angry
A_S4	118645	113698	116171.5	2.4826	-0.7528	1.14E+05	201.3993	5.83E+04	241.4645	90	Angry
A_S5	116238	111530	113884	2.3674	-0.0038	1.12E+05	309.2317	1.43E+05	378.4524	70	Angry
A_S6	117863	112251	115057	2.2554	0.3295	1.13E+05	311.1568	1.39E+05	373.331	80	Angry
A_S7	118472	114448	116460	3.1061	-0.3591	1.08E+05	142.3957	3.24E+04	180.041	74	Angry
A_S8	118607	114334	116470	1.8625	0.1661	1.18E+05	269.0124	9.36E+04	305.93	92	Angry
A_S9	118556	112509	115532	1.9988	-0.1346	1.18E+05	209.9363	5.92E+04	243.345	81	Angry
A_S10	119568	113053	116310	1.953	-0.448	1.19E+05	327.9787	1.39E+05	372.5859	102	Angry
H_S1	111584	110368	1.11E+05	2.9886	0.4422	1.11E+05	185.5407	5.36E+04	231.5724	76	Happy
H_S2	112408	110280	111344	2.2368	0.6053	1.11E+05	430.039	2.59E+05	509.3132	78	Happy
H_S3	114304	112378	113341	2.2199	-0.4183	1.16E+05	299.4117	1.23E+05	350.4022	66	Happy
H_S4	112575	108713	110644	11.3788	2.56	1.09E+05	189.7319	9.08E+04	301.3572	78	Happy
H_S5	111369	111047	1.11E+05	2.7524	0.4741	1.20E+05	425.7182	2.92E+05	540.5537	64	Happy
H_S6	114187	113263	1.14E+05	2.4314	0.6468	1.15E+05	707.4147	7.14E+05	844.8233	72	Happy
H_S7	114704	114063	114383	2.7898	0.6799	1.18E+05	524.9938	4.45E+05	667.2781	63	Happy
H_S8	114153	112795	114474	2.8637	0.7127	1.15E+05	942.3012	1.46E+06	1.21E+03	80	Happy
H_S9	113129	112388	1.13E+05	3.849	-0.9285	1.16E+05	357.9615	2.14E+05	462.5541	71	Happy
H_S10	112126	111066	1.12E+05	2.8947	0.434	1.17E+05	316.6883	1.55E+05	393.8948	90	Happy
N_S1	115839	114839	1.15E+05	2.0222	0.0173	1.15E+05	195.8882	5.33E+04	230.8929	63	Neutral
N_S2	114723	112677	1.14E+05	2.0686	0.5344	1.13E+05	464.026	2.92E+05	540.4605	70	Neutral
N_S3	114149	111275	1.13E+05	2.7813	-0.5988	1.13E+05	510.738	4.13E+05	642.9525	57	Neutral

Figure 62: Excel sheet filled with the features.

The response was also added next to each subject. After the completion of the excel sheet, it was imported to the Matlab software as a table. Next the workspace was introduced to the classification learner. Cross validation was selected with a 2-fold validation selected. The system was then trained with the available classifiers. The classifier with the most accuracy was the complex tree and the bagged tree classifier as it got a 77.5% accuracy.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

6- Pre-processing

6.1- Tweets:

Having the corpus and the datasets ready with all resources gathered, we can proceed with the preprocessing, the preprocessing phase is the most important phase, it will affect the results at the end, if the preprocessing is good, the results will be highly accurate and precise. As a result, we will have the dataset ready for training and testing.

The preprocessing will be divided into several phases:

- Replace all the emoticons with their sentiment tag which is pos or neg using the resource provided.
- Replace all URLs with a tag ||url||.
- Remove Unicode characters.
- Decode HTML entities.
- Reduce all letters to lowercase.
- Replace all usernames/targets @ with ||target||.
- Replace all acronyms with their translation.
- Replace all negations by tag ||not||.
- Replace a sequence of repeated characters by two characters.

Not to forget the importance of the tagging and lexicons phases inside the preprocessing we are doing here. In addition, those two are extremely important in terms of NLP and Sentiment Analysis. All these techniques will help us reach better results in the validation and testing phases.

All the tweets in both datasets are pre-processed with the same steps and sequence.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

order.

6.1.1- Emoticons

We replace all the emoticons provided in the tweets, with their neg or pos polarities presented in the resource data dictionary. Thus, we will be using some search functions to search for the words in the dictionary and another function to replace the word with the polarity. We will be showing some pictures of dataset2 since same procedure is being performed for both datasets.

Out[6]:

	Sentiment	SentimentText
0	0	is so sad for my APL friend.....
1	0	I missed the New Moon trailer...
2	1	omg its already 7:30 :O
3	0	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
4	0	i think mi bf is cheating on me!!! T_T
5	0	or i just worry too much?
6	1	Juuuuuuuuuuuuuuuussssst Chillin!!
7	0	Sunny Again Work Tomorrow :-| TV Tonight
8	1	handed in my uniform today . i miss you already
9	1	hmmmm.... i wonder how she my number @-)

Figure 63: Tweets Before replacing emoticons.

0]:

	Sentiment	SentimentText
0	0	is so sad for my APL friend.....
1	0	I missed the New Moon trailer...
2	1	omg its already 7:30 ||pos||
3	0	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
4	0	i think mi bf is cheating on me!!! ||neg||
5	0	or i just worry too much?
6	1	Juuuuuuuuuuuuuuuussssst Chillin!!
7	0	Sunny Again Work Tomorrow ||neg|| TV Tonight
8	1	handed in my uniform today . i miss you already
9	1	hmmmm.... i wonder how she my number ||pos||

Figure 64: Tweets after replacing emoticons.

This dataset contains 19469 positive emoticons and 11025 negative ones.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

6.1.2- Websites or URLS

All websites or URL links will be removed, we will substitute them with their corresponding tag. We recorded almost 739824 URLS in the seconds dataset, so we perform the same substitution done with emoticons here.

	Sentiment	SentimentText
50	0	baddest day eveer.
51	1	bathroom is clean..... now on to more enjoyable tasks.....
52	1	boom boom pow
53	0	but i'm proud.
54	0	congrats to helio though
55	0	David must be hospitalized for five days end of July (palatine tonsils). I will probably never see Katie in concert.
56	0	friends are leaving me 'cause of this stupid love http://bit.ly/ZoxZC
57	1	go give ur mom a hug right now. http://bit.ly/azFwv
58	1	Going To See Harry Sunday Happiness
59	0	Hand quilting it is then...

Figure 65: Tweets before replacing URLS.

	Sentiment	SentimentText
50	0	baddest day eveer.
51	1	bathroom is clean..... now on to more enjoyable tasks.....
52	1	boom boom pow
53	0	but i'm proud.
54	0	congrats to helio though
55	0	David must be hospitalized for five days end of July (palatine tonsils). I will probably never see Katie in concert.
56	0	friends are leaving me 'cause of this stupid love ‖url‖
57	1	go give ur mom a hug right now. ‖url‖
58	1	Going To See Harry Sunday Happiness
59	0	Hand quilting it is then...

Figure 66: Tweets after replacing URLS.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

6.1.3- Remove Unicode Characters

Now we remove all Unicode characters from our dataset, and we only keep the ASCII ones, because Unicode characters will cause problems during the tokenization process.

	Sentiment	SentimentText
1578592	1	'Zu SpÃ¤t' by Die Ã„rzte. One of the best bands ever
1578593	1	Zuma bitch tomorrow. Have a wonderful night everyone goodnight.
1578594	0	zummie's couch tour was amazing....to bad i had to leave early
1578595	0	ZuneHD looks great! OLED screen @720p, HDMI, only issue is that I have an iPhone and 2 iPods . MAKE IT A PHONE and ill buy it @micro...
1578596	1	zup there ! learning a new magic trick
1578597	1	zyklonic showers *evil*
1578598	1	ZZ Top â€' I Thank You ...@hawaiibuzz .....Thanks for your music and for your ear(s) ...ALL !!!! Have a fab... â™« ||url||
1578599	0	zzz time. Just wish my love could B nxt 2 me
1578600	1	zzz twitter. good day today. got a lot accomplished. imstorm. got into it w yet another girl. dress shopping tmrw
1578601	1	zzz's time, goodnight. ||url||

Figure 67: Tweets before removing Unicode characters.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

	Sentiment	SentimentText
1578592	1	'Zu Spt' by Die rzte. One of the best bands ever
1578593	1	Zuma bitch tomorrow. Have a wonderful night everyone goodnight.
1578594	0	zummie's couch tour was amazing....to bad i had to leave early
1578595	0	ZuneHD looks great! OLED screen @720p, HDMI, only issue is that I have an iPhone and 2 iPods . MAKE IT A PHONE and ill buy it @micro...
1578596	1	zup there ! learning a new magic trick
1578597	1	zyklonic showers *evil*
1578598	1	ZZ Top I Thank You ...@hawaiibuzz .....Thanks for your music and for your ear(s) ...ALL !!!! Have a fab... ||url||
1578599	0	zzz time. Just wish my love could B nxt 2 me
1578600	1	zzz twitter. good day today. got a lot accomplished. imstorm. got into it w yet another girl. dress shopping tmrw
1578601	1	zzz's time, goodnight. ||url||

Figure 68: Tweets after removing Unicode characters.

6.1.4- Decoding HTML Entities

Now we will decode some HTML entities, these are considered a problem in text classification projects.

```
'   Cannot get chatroom feature to work.   Updated Java to 10, checked ports, etc. I can see video, but in the &quot;chat,&qu
ot; only a spinning circle.   '
```

Figure 69: Tweets before decoding HTML Entities.

```
'   Cannot get chatroom feature to work.   Updated Java to 10, checked ports, etc. I can see video, but in the "chat," only a
spinning circle.   '
```

Figure 70: Tweets after decoding HTML Entities.

6.1.5- Reducing all letters to lower case.

This part is very easy, we should simply reduce all letters to lower case, this will make it easier for all

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

dictionaries to work properly especially stop words.

	Sentiment	SentimentText
0	0	is so sad for my APL friend.....
1	0	I missed the New Moon trailer...
2	1	omg its already 7:30 ||pos||
3	0	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
4	0	i think mi bf is cheating on me!!! ||neg||
5	0	or i just worry too much?
6	1	Juuuuuuuuuuuuuuuussssst Chillin!!
7	0	Sunny Again Work Tomorrow ||neg|| TV Tonight
8	1	handed in my uniform today . i miss you already
9	1	hmmmm.... i wonder how she my number ||pos||

Figure 71: Tweets before reducing the letters to lower case.

	Sentiment	SentimentText
0	0	is so sad for my apl friend.....
1	0	i missed the new moon trailer...
2	1	omg its already 7:30 ||pos||
3	0	.. omgaga. im sooo im gunna cry. i've been at this dentist since 11.. i was suposed 2 just get a crown put on (30mins)...
4	0	i think mi bf is cheating on me!!! ||neg||
5	0	or i just worry too much?
6	1	juuuuuuuuuuuuuuuussssst chillin!!
7	0	sunny again work tomorrow ||neg|| tv tonight
8	1	handed in my uniform today . i miss you already
9	1	hmmmm.... i wonder how she my number ||pos||

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Figure 72: Tweets after reducing the letters to lower case.

6.1.6 – Replacing all usernames.

Since we do not need to consider the usernames in the sentiment detection phase, we simply replace them with the tag ||target||. Knowing that in dataset2 we have 735757 mentions of usernames.

	Sentiment	SentimentText
45	1	@ginaaa <3 go to the show tonight
46	0	@spiral_galaxy @ymptweet it really makes me sad when i look at muslims reality now
47	0	- all time low shall be my motivation for the rest of the week.
48	0	and the entertainment is over, someone complained properly.. @ruptureapture experimental you say? he should experiment with a me...
49	0	another year of lakers .. that's neither magic nor fun ...
50	0	baddest day eveer.
51	1	bathroom is clean..... now on to more enjoyable tasks.....
52	1	boom boom pow
53	0	but i'm proud.
54	0	congrats to helio though

Figure 73: Tweets before replacing the usernames.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

	Sentiment	SentimentText
45	1	||target|| <3 go to the show tonight
46	0	||target|| ||target|| it really makes me sad when i look at muslims reality now
47	0	- all time low shall be my motivation for the rest of the week.
48	0	and the entertainment is over, someone complained properly.. ||target|| experimental you say? he should experiment with a melody...
49	0	another year of lakers .. that's neither magic nor fun ...
50	0	baddest day eveer.
51	1	bathroom is clean..... now on to more enjoyable tasks.....
52	1	boom boom pow
53	0	but i'm proud.
54	0	congrats to helio though

Figure 74: Tweets after replacing the usernames.

6.1.7- Acronyms

All acronyms will be substituted with their corresponding translation provided in the acronym dictionary, thus tokenizing the tweets with removing all punctuations and splitting the data in texts. NLTK library will be used even though it is considered slower in pace, but it is more accurate, but it is not a severe problem). Even though replacements will not be perfect, a simple example is the acronym "I'm" meaning "instant message". It would not be surprising that in most of the cases, "I'm" means "I am". For that, some improvements will be done later to enhance our results.

```
[('lol', 59000),
 ('u', 54557),
 ('im', 51099),
 ('2', 42645),
 ('gonna', 23716),
 ('4', 18610),
 ('dont', 18363),
 ('wanna', 16357),
 ('ok', 16104),
 ('ur', 12960),
 ('omg', 12178),
 ('n', 10415),
 ('ya', 9948),
 ('gotta', 9243),
 ('r', 8132),
 ('tho', 7696),
 ('tv', 6246),
 ('o', 6002),
 ('kinda', 5953),
 ('pic', 5945)]
```

Figure 75: Count of acronyms in dataset 2

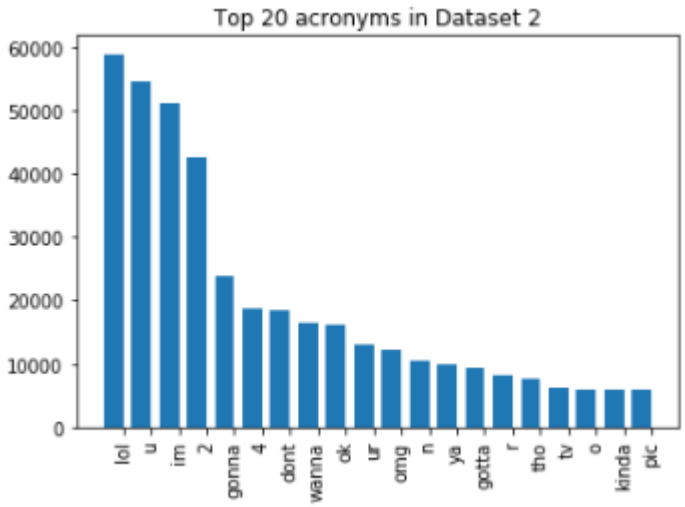


Figure 76: Top 20 acronyms in dataset 2

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

6.1.8- Replace all negations.

Using the negation dictionary, we will be replacing all negations with ||not|| knowing that when we encounter a negation, if a positive or a negative word is followed by it, the word or phrase will be opposed, positive become negative and vice versa, this will be done when finding positive words as well as the negative ones.

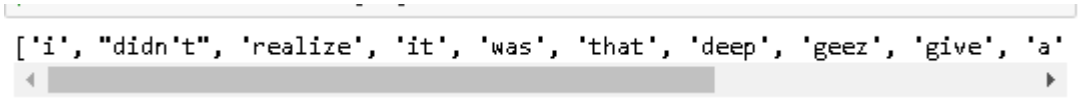


Figure 77: Tweets before replacing negations.

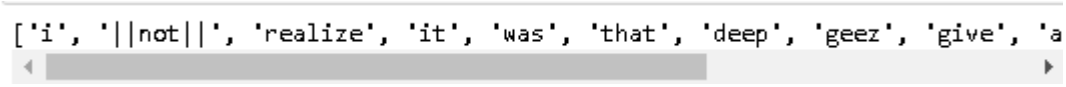


Figure 78: Tweets after replacing negations.

6.1.9- Replace repeated characters.

There are many words that might contain repeated characters, we should reduce the number of these.

characters to reduce the feature space to be used later.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

	Sentiment	SentimentText
1578604	1	[zzzz, no, work, tomorrow, yayyy]
1578605	1	[zzzzz, time, tomorrow, will, be, a, busy, day, for, serving, loving, people, love, you, all]
1578606	0	[zzzzz, want, to, sleep, but, at, sister's, in, laws's, house]
1578607	1	[zzzzzz, finally, night, tweeters]
1578608	1	[zzzzzzz, sleep, well, people]
1578609	0	[zzzzzzzzzz, wait, no, i, have, homework]
1578610	0	[zzzzzzzzzzzzz, whatever, what, am, i, doing, up, again]
1578611	0	[zzzzzzzzzzzzzzzzzzz, i, wish]

Figure 79: Tweets before replacing repeated characters.

	Sentiment	SentimentText
1578604	1	[zz, no, work, tomorrow, yayy]
1578605	1	[zz, time, tomorrow, will, be, a, busy, day, for, serving, loving, people, love, you, all]
1578606	0	[zz, want, to, sleep, but, at, sister's, in, laws's, house]
1578607	1	[zz, finally, night, tweeters]
1578608	1	[zz, sleep, well, people]
1578609	0	[zz, wait, no, i, have, homework]
1578610	0	[zz, whatever, what, am, i, doing, up, again]
1578611	0	[zz, i, wish]

Figure 80: Tweets after replacing repeated characters.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

7- Machine Learning

[5] [6] [7]

7.1 – Tweets

7.1.1- Procedure

After finishing up with all the preprocessing techniques, we focus on the machine learning part. We have 3 important methods to take into consideration when dealing with text classification projects. Naïve bayes, SVM and N-Grams. The last two methods are most used, but with large datasets SVM cannot function, that is why we did two datasets, 1 large dataset and the other is a sample of it to test SVM and other algorithms with it.

First we will be using most of the Scikit-learn tools **[21][22][23][24][25][26][27][28][29]** we will start by splitting our first dataset into 90% training and 10% for testing since our dataset contains 10,000 records so we take 9000 for training and 1000 for testing (this dataset is small hence we will be performing Logistic Regression **[25]** K-Neighbors **[26]** SVM also known as SVC **[27]** SGD **[24]** Decision tree classifier **[22]** and finally, Multinomial Naïve bayes **[23]**. Whereas for dataset2 which is the large dataset, we will split it into 75% for training and 25% for testing, knowing that we have over 2million data records in this dataset. We will be only performing the Multinomial naïve bayes algorithm because the rest cannot work in such a massive dataset. And at the end based on our evaluation and scores metrics we will choose the best model and we will be performing the testing on new data.

We used K-folding method with the K=10 folds this is not a basic value to take, but there is no specific metrics to detect this value, but usually K is set to 10. However, that means for example in dataset 1, we will have 10 folds, so since the data is 10,000 then 10,000/10=1000 that means, we have a validation data size of 1000 each fold, we perform the validation and training, and at the end we average the results to get the most approximate scores. We also used the Count vectorizer function **[21]** which is a very famous function to be used in text classification projects, it is basically used for feature selection and extraction, it follows the bag of words concept, it represents the frequency of every single word in the dataset, and takes it into account as

a feature, to be used later using the classification techniques and algorithms. However, N-Grams is considered another milestone we must reach and take advantage of, the count vectorizer function operates usually at Ngram=(1,1) which is unigrams, so it takes word by word, but there is also unigrams (2,2) and (1,2) unigrams and bigrams, every single one might impact the results either positively or negatively depends, unigrams and bigrams are commonly known as the best approach, we will be trying all of them in this project, to have a better vision towards the results, and t choose wisely the best model.

We will be also using the Pickle library **[44]** to save each model as a PKL file, along with its count vectorizer function that contains the N-gram parameter, so that later for testing, we do not need to rerun the model, we can only call it and test on unseen tweets. So, we will be saving the models with the vectorizer function in each step. But we did a simple trick in the code, so since the code will be looping through each fold and calculating the metrics, the pickle function will be saving the fold with the best F1 score, so if we choose this model later on for testing, and we try to call the pkl file, it will give us like 1 to 4% better F1 score and accuracy, because we are calling the best fold among the 10folds of the best model we chose, and that's a plus for us in testing. At the end we will be calculating the overall precision, accuracy, F1 scores, recall, confusion matrix of all models implemented with their corresponding characteristics.

In addition to that, we will be performing all models before and after removing stop words and stemming, because stop words are considered irrelevant to text classification tasks and might affect the results, same goes to stemming in NLP, we used NLTK library for stemming **[33][34]** and the stop words dataset we got with some python function for removing them**.** We will visualize the results before and after these two steps with all algorithms and all possible values of the N-grams, in both datasets, the more possibilities we have, the more we reach the most optimal model for testing, and that is what we aim towards. The results will be shown later in this report.

7.1.2- Evaluation and Metrics

[28][29]

7.1.2.1- First Dataset

We will begin with the first dataset that has over 10,000 data records, which is applicable for the classifiers we will use in this procedure (Decision tree, logistic regression, SGD, SVM, K-Neighbors) with the naïve

bayes algorithm.

7.1.2.1.1- Before removing stop words and stemming:

We will be starting with a table that shows results we got for dataset 1 using different classifiers before removing stop words and stemming and using **only unigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N.B	70.614%	71.333%	71.014%	68.056%	[[3376 1124] [1429 3071]]
Decision Tree	67.667%	67.778%	67.580%	66.667%	[[6370 2630] [3055 5945]]
K-Neighbors	66.686%	62.222%	64.258%	67.695%	[[8841 4659] [4430 9070]]
SGD	67.894%	72.111%	72.668%	72.826%	[[11957 6043] [5652 12348]]
SVM	69.022%	72%	71.175%	72.260%	[[15110 7390] [6751 15749]]
Logistic Regression	69.675%	72.444%	70.127%	75.570%	[[18271 8729] [7899 19101]]

Figure 81:Table showing Dataset 1 Metrics for Unigrams only before stemming and removing stop words.

Now we will perform the same procedure but with **Unigrams and Bigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N.B	70.931%	71.222%	76.528%	65.756%	[[3495 1005] [1474 3026]]
Decision Tree	68.426%	64.667%	67.329%	64.210%	[[6484 2516] [3018 5982]]
K-Neighbors	65.816%	59.444%	59.432%	63.973%	[[9032 4468] [4707 8793]]
SGD	67.502%	72.889%	70.021%	76.782%	[[12126 5874] [5840 12160]]
SVM	68.842%	72%	67.427%	77.381%	[[15184 7316] [6833 15667]]
Logistic Regression	69.669%	76.222%	76.279%	74.545%	[[18403 8597] [7952 19048]]

Figure 82: Table showing Dataset 1 Metrics for Unigrams and Bigrams only before stemming and removing stop words.

Now with **Bigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Multinomial N.B	68.359%	67.555%	69.544%	63.736%	[[3283 1217] [1530 2970]]
Decision Tree	65.445%	62.444%	68.591%	59.520%	[[6274 2726] [3292 5708]]
K-Neighbors	62.305%	54.778%	57.350%	57.950%	[[8800 4700] [5268 8232]]
SGD	64.160%	66.667%	64.672%	74.115%	[[11710 6290] [6508 11492]]
SVM	65.606%	66.778%	61.565%	81.944%	[[14159 8341] [7371 15129]]
Logistic Regression	66.444%	68.778%	65.690%	72.854%	[[17136 9864] [8582 18418]]

Figure 83:Table showing Dataset 1 Metrics for Bigrams only before stemming and removing stop words.

7.1.2.1.2- After removing stop words:

Now we will see after removing stop words with **unigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N.B	71.038%	71.444%	74.823%	67.949%	[[3396 1104] [1412 3088]]
Decision Tree	68.082%	68%	66.988%	64.802%	[[6403 2597] [3017 5983]]
K-Neighbors	67.014%	65.667%	63.019%	74.720%	[[8854 4646] [4371 9129]]
SGD	68.139%	70%	71.938%	69.164%	[[11983 6017] [5601 12399]]
SVM	69.232%	73.222%	69.215%	78.454%	[[15158 7342] [6709 15791]]
Logistic Regression	69.883%	74.444%	75.862%	74.894%	[[18318 8682] [7841 19159]]

Figure 84: Table showing Dataset 1 Metrics for Unigrams only after removing stop words.

Now we move to **Unigrams and Bigrams as well.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Multinomial N.B	70.960%	71.444%	70.844%	65.952%	[[3494 1006] [1470 3030]]
Decision Tree	67.992%	67.111%	67.035%	67.333%	[[6488 2512] [3075 5925]]
K-Neighbors	65.312%	56.889%	57.720%	69.118%	[[8976 4524] [4783 8717]]
SGD	67.010%	70.556%	69.058%	70.804%	[[12171 5829] [6005 11995]]
SVM	68.440%	71.444%	68%	75.467%	[[15214 7286] [6991 15509]]
Logistic Regression	69.386%	75.222%	74.894%	77.024%	[[18459 8541] [8102 18898]]

Figure 85: Table showing Dataset 1 Metrics for Unigrams and bigrams only after removing stop words.

Now **only Bigrams.**

	F1 Score	**Accuracy**	**Precision**	**Recall**	**Confusion Matrix**
Multinomial N.B	67.960%	66.889%	72.794%	61.363%	[[3260 1240] [1545 2955]]
Decision Tree	65.146%	61.889%	62.413%	62.413%	[[6202 2798] [3300 5700]]
K-Neighbors	62.062%	55.222%	55.344%	52.009%	[[8704 4796] [5267 8233]]
SGD	63.880%	66.111%	63.944%	72.135%	[[11606 6394] [6530 11470]]
SVM	65.362%	67.444%	64.634%	80.478%	[[14064 8436] [7404 15096]]
Logistic Regression	66.258%	70%	68.442%	74.222%	[[17034 9966] [8604 18396]]

Figure 86: Table showing Dataset 1 Metrics for bigrams only after removing stop words.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

7.1.2.1.3- After Stemming:

We will start with the **unigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N.B	71.269%	71.333%	71.635%	68.037%	[[3415 1085] [1407 3093]]
Decision Tree	68.320%	64.222%	63.393%	63.392%	[[6458 2542] [3013 5987]]
K-Neighbors	67.032%	62.222%	61.847%	67.248%	[[8892 4608] [4389 9111]]
SGD	68.173%	72.444%	76.267%	76.267%	[[12045 5955] [5627 12373]]
SVM	69.258%	73.777%	72.921%	75.831%	[[15215 7285] [6730 15770]]
Logistic Regression	69.864%	72.889%	70.876%	77.506%	[[18363 8637] [7874 19126]]

Figure 87: Table showing Dataset 1 Metrics for unigrams only after stemming.

Now we will proceed with **unigrams and bigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N. B	70.975%	73.667%	77.975%	67.249%	[[3519 981] [1482 3016]]
Decision Tree	68.340%	68.667%	68.065%	66.819%	[[6568 2432] [3072 5928]]
K-Neighbors	65.551%	56.556%	55.556%	59.184%	[[9093 4407] [4791 8709]]
SGD	67.438%	70.222%	67.194%	70.714%	[[12231 5769] [5911 12089]]
SVM	68.765%	70.556%	68.303%	75.225%	[[15268 7232] [6901 15599]]
Logistic Regression	69.633%	73.667%	71.522%	75.632%	[[18496 8504] [8013 18987]]

Figure 88: Table showing Dataset 1 Metrics for unigrams and bigrams only after stemming.

Finally with **Bigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial N.B	68.426%	72.222%	73.077%	68.778%	[[3294 1206] [1531 2969]]
Decision Tree	65.598%	62.556%	65.185%	57.391%	[[6240 2760] [3259 5741]]
K-Neighbors	62.114%	54.889%	56.444%	64.496%	[[8732 4768] [5264 8236]]
SGD	63.945%	68.778%	68.826%	72.805%	[[11625 6375] [6515 11485]]
SVM	65.406%	70.444%	66.608%	83.553%	[[14080 8420] [7393 15107]]
Logistic Regression	66.184%	69.333%	71.097%	70.798%	[[17035 9965] [8632 18368]]

Figure 89: Table showing Dataset 1 Metrics for bigrams only after stemming.

7.1.2.2- Second Dataset

We will start with the second dataset that contains approximately 1.6 million data records and we will be performing only multinomial naïve bayes classifier in this dataset.

5.1.2.2.1- Before removing stop, words and stemming:

We will be starting with a table that shows the results we got for dataset2 using Multinomial N.B before stemming and removing stop words and using **only unigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	77.697%	77.746%	78.298%	76.994%	[[465029 126297] [135903 456729]]

Figure 90: Table showing Dataset 2 Metrics for Unigrams only before stemming and removing stop words.

Now we will see the results but **using Unigrams and bigrams as well.**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	79.634%	80.582%	81.750%	77.689$	[[486614 104712] [131265 461367]]

Figure 91:Table showing Dataset 2 Metrics for Unigrams and Bigrams only before stemming and removing stop words.

Now we will see the results **using only Bigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	78.567%	79.111%	80.616%	76.857$	[[480588 110738] [137546 455086]]

Figure 92: Table showing Dataset 2 Metrics for Bigrams only before stemming and removing stop words.

7.1.2.2.2- After removing stop words:

We will start by viewing the results after removing stop words **for unigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	73.574%	72.222%	74.473%	69.281$	[[3347 1075] [1288 3290]]

Figure 93: Table showing Dataset 2 Metrics for Unigrams only after removing stop words.

We can notice that the confusion matrix values has decreased massively, yet the difference between the false and true ones is still the same, and the scores has decreased, which means it might be a bad step to be done so far.

We will check it now **with bigrams and unigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	79.630%	80.005%	81.380%	77.724$	[[486558 104768] [131271 461361]]

Figure 94: Table showing Dataset 2 Metrics for Unigrams and Bigrams only after removing stop words.

We can notice that the scores have increased with unigrams and bigrams, even though it is still slightly less than the performance of unigrams and bigrams before removing stop words.

Now we will check it with **bigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	78.569%	78.867%	80.189%	76.632%	[[480586 110740] [137532 455100]]

Figure 95:Table showing Dataset 2 Metrics for Bigrams only after removing stop words.

We can notice that the results decreased a little, knowing that the bigrams only are the weakest N-gram parameter to be taken.

7.1.2.2.3- After Stemming:

[33]

We will start with the **unigrams only** after stemming.

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	78.569%	78.867%	80.189%	76.632%	[[480586 110740] [137532 455100]]

Figure 96:Table showing Dataset 2 Metrics for Unigrams only after stemming.

Now we will try with **Unigrams and Bigrams.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	79.622%	79.962%	81.474%	77.631%	[[486494 104832] [131305 461327]]

Figure 97: Table showing Dataset 2 Metrics for Unigrams and Bigrams only after stemming.

And finally with **Bigrams only.**

	F1 Score	Accuracy	Precision	Recall	Confusion Matrix
Multinomial Naïve Bayes	78.575%	79.122%	80.616%	76.745%	[[480612 110714] [137494 455138]]

Figure 98: Table showing Dataset 2 Metrics for Bigrams only after stemming.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

7.1.3- Choosing the best model:

Since we are done with all possible models, now it is time to choose the best model. We have different metrics like accuracy, precision, recall and F1 score, which one should we give a priority in our case? Well, basically accuracy and F1 score are the most important, accuracy is important when the TP and TN values are very essential in our application, whereas for the F1 score, it is considered very essential when the FN and FP are pretty much crucial and important, in our text classification case, the FN and FP values are more important, and not to forget that in most classification applications, there is an always imbalanced distribution, so as a result, the F1-score is the most important.

So, this **table below**, which is from **figure 88,** it corresponds to **the Multinomial Naïve Bayes algorithm** used in **dataset 2** which has almost **1.5million data records**, it is basically used with **both unigrams and bigrams before stemming and removing stop words**. This model has gained the maximum F1 score among all models, in addition to that, its accuracy and precision values are also among the top 2 to 3 models we have seen. Thus, we will be choosing this algorithm for our testing procedure.

	F1 Score	**Accuracy**	**Precision**	**Recall**	**Confusion Matrix**
Multinomial Naïve Bayes	79.634%	80.582%	81.750%	77.689$	[[486614 104712] [131265 461367]]

However, this F1 score value beats most of the reports and latest research regarding text classification mentioned earlier in the Sources section. In addition to that, notice that this score is the average score computed on all the folds of the model, we used 10 folds, and we averaged them, but as we mentioned earlier, we used the pickle library to save the models, but we intended to use this library in a way that it will save the best fold and not the whole folds, and that's a plus for us, so basically the model that we will be using in the testing phase, its F1 score and accuracy will be 1 or 2% higher than the one we have here and that's a plus for us as well.

7.1.4- Testing:

After we chose the best model for text classification, we will perform some testing. The thing is that we need to acquire data from twitter, to do that we need to download Tweepy library **[30]** which is a library presented in python, it is used so that users can access Twitter API, and thus using twitter data in their projects and future work. It is also used in a variety of applications used in text classification and processing of twitter data. We will also need Access tokens **[45],** these tokens are needed for a user to receive access to twitter data. Certain questions will be asked, and an application must be filled online in an official twitter website for credentials and API. The user shall receive access with necessary information and tokens needed to write them inside the python code, so that the user gets full access to twitter data. After receiving the access, user can use the tweepy functions inside python to retrieve tweets for any user on twitter application and start analyzing them, along with several other capabilities to be done as well.

We start by choosing a twitter user to analyze his/her tweets, we first use pandas and tweepy **[39][40][41][30][45]** to retrieve and view the new tweets of the user we want to test our model to. After that, we can choose the number of tweets to be retrieved, in this case the api.user_timeline function **[30]** presented in the tweepy library can only retrieve a maximum of 200 tweets, it can be solved through some a little complex coding method, but it is not our purpose for this project, 200 is enough for us as a maximum number of tweets to be retrieved.

As we can see in **Figure 99,** we retrieved the first 200 tweets from President Joe Biden's account (Date: 22/4/2021 3:19AM Beirut time zone) we notice that the tweets presented are un preprocessed and cannot be dealt with properly. We represent the new data as a dataset using Pandas library **[39][40][41]**

```
Type the twitter username that you want to analyze his or her tweets: joebiden
How many tweets do you want to extract? 200
```

Out[9]:

	Tweets
0	RT @POTUS: The guilty verdict does not bring back George Floyd. But through the family's pain, they are finding purpose so George's legacy...
1	RT @POTUS: Today, a jury in Minnesota found former Minneapolis Police Officer Derek Chauvin guilty of murdering George Floyd.\n \nThe verdict...
2	RT @WhiteHouse: Live: President Biden and Vice President Harris address the nation on the verdict in the trial of Derek Chauvin. https://t....
3	If we act now on the American Jobs Plan, in 50 years, people will look back and say this was the moment that America won the future. https://t.co/ZOJGr9E2i3
4	Today, every adult is eligible to get a COVID-19 vaccine. Better days are ahead.
...	...
195	I urge every American to:\n\n- Wear a mask\n- Stay socially distanced\n- Avoid large indoor gatherings\n\nWe can save countless lives if we step up together.
196	Getting America vaccinated will be one of the greatest operational challenges we've ever faced, but my administration will spare no effort to get it done.\n\nWe're going to ensure the vaccine is distributed quickly, equitably, and free of charge to every American.
197	Folks, I just received the second dose of my COVID-19 vaccine — and just like the first dose, it was safe, quick, and painless. \n\nI urge everyone to get vaccinated once it's your turn. Because only together can we save lives and beat this virus. https://t.co/w1m8gEh2iL
198	The work of the next four years must be the restoration of democracy and the recovery of respect for the rule of law, and the renewal of a politics that's about solving problems — not stoking the flames of hate and chaos.
199	I'm asking Ambassador Bill Burns to lead the Central Intelligence Agency because he's dealt with many of the thorniest global challenges we face. As a legendary career diplomat, he approached complex issues with honesty, integrity and skill. That's exactly how he'll lead the CIA. https://t.co/ypnuH016BV

200 rows × 1 columns

Figure 99: Extracting A Twitter user tweets (Joe Biden) to perform the predictions.

Since we chose our model which was for dataset2 the multinomial naïve bayes with unigrams and bigrams before stemming and preprocessing, thus when we perform testing on new data, the new data should be 100% identical to the data we used in our model, otherwise the model won't work properly, because it was trained based on those preprocessed data, and it will be unable to give good predictions on new data if we don't have any preprocessing on it, because the testing data should be exactly the same as the training data used before, this is a fundamental rule of machine learning**[5][6][7]**. This means that the new data should be preprocessed with all the steps we did to our model before we implement the classifiers, as a result the new dataset will be preprocessed with all steps presented in **Section 3.1 (Preprocessing of twitter data).**

After performing all preprocessing techniques used before to our new data, which is Joe Biden's tweets, we can notice in **figure 100** the new tweets after preprocessing.

Out[10]:

	Tweets
0	‖target‖ the guilty verdict does ‖not‖ bring back george floyd but through the familys pain they are finding purpose so georges legacy
1	‖target‖ today a jury in minnesota found former minneapolis police officer derek chauvin guilty of murdering george floyd the verdict
2	‖target‖ live president biden and vice president harris address the nation on the verdict in the trial of derek chauvin ‖url‖
3	if we act now on the american jobs plan in years people will look back and say this was the moment that america won the future ‖url‖
4	today every adult is eligible to get a covid vaccine better days are ahead
...	...
195	i urge every american to wear a mask stay socially distanced avoid large indoor gatherings we can save countless lives if we step up together
196	getting america vaccinated will be one of the greatest operational challenges what ever ever faced but my administration will spare no effort to get it done were going to ensure the vaccine is distributed quickly equitably and free of charge to every american
197	folks i just received the second dose of my covid vaccine and just like the first dose it was safe quick and painless i urge everyone to get vaccinated once its your turn because only together can we save lives and beat this virus ‖url‖
198	the work of the next four years must be the restoration of democracy and the recovery of respect for the rule of law and the renewal of a politics thats about solving problems ‖not‖ stoking the angry comments of hate and chaos
199	instant message asking ambassador bill burns to lead the central intelligence agency because hes dealt with many of the thorniest global challenges we face as a legendary career diplomat he approached complex issues with honesty integrity and skill thats exactly how hell lead the cia ‖url‖

Figure 100: Joe Biden's tweets (new data) after preprocessing

We should not perform neither stemming nor remove stop words, because our model was chosen before removing stop words and stemming, and thus the new data should be the same as well.

After we prepared our new data for testing, we should not forget that the model we chose was with unigrams and bigrams, so the count vectorizer function that will performed to these new tweets should be with parameter (1,2) which is for unigrams and bigrams as well, but luckily for us, using the Pickle library **[44]** we saved the model along with its count vectorizer function and parameter using the library as PKL file, so all we need to do is just load the model with the function, and we transform our new tweets using the count vectorizer into a bag of words, and then we predict using the model, and that's all.
Let us check the most recent words Joe Biden is using in his recent 200 tweets using the WordCloud library **[52]** presented in **figure 101**.

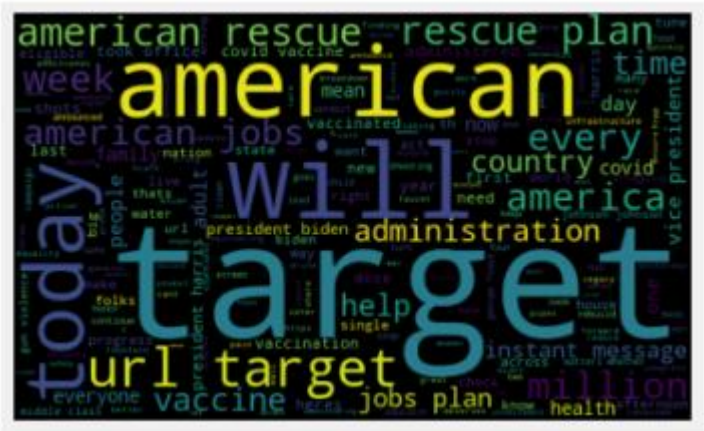


Figure 101: Most used words in Joe Biden's tweets

Let us take jobs as an example, so we take something specific to use in our testing, so now we create a new dataset that has only the tweets that are related to jobs posted by Joe Biden. Let us see in **figure 102.**

```
Type a word that you want your tweets to be related to: jobs
```

Out[40]:

	Tweets
0	if we act now on the american jobs plan in years people will look back and say this was the moment that america won the future ‖url‖
1	‖target‖ imagine a future where we lead the world and tackle the threat of climate change with american jobs and ingenuity we can make
2	‖target‖ clean drinking water is infrastructure the american jobs plan will improve our water infrastructure by replacing of t
3	‖target‖ this week i announced the american jobs plan which will invest in our infrastructure and strengthen americas competitiveness
4	‖target‖ the american jobs plan will ensure every american can turn on the faucet or fountain and drink clean water ‖url‖
5	the american jobs plan is the largest american jobs investment since world war ii
6	under the american jobs plan of our nations lead pipes and service lines will be replacedso every child in america can turn on the faucet or fountain and drink clean water we ‖not‖ delay another minute
7	wall street didnt build this countrythe great american middle class did this time when we rebuild the middle class under the american jobs plan were bringing everyone along
8	delivering for the american people is what the american rescue plan was about its what the american jobs plan is about we can do this we have to do this we will do this
9	the american jobs plan is a once in a generation investment in america it will modernize miles of highways roads and main streets repair bridges desperately in need of upgrades replace of our nations lead pipes and service lines
10	millions of americans lost their jobs last year heres the truth we all do better when we all do well its time to build our economy from the bottom upand the middle outnot the top down
11	we ‖not‖ only have an economic imperative to act now we have a moral obligation in this pandemic in america we ‖not‖ let people go hungry we ‖not‖ let people get evicted we ‖not‖ watch nurses and educators lose their jobs we must act

Figure 102: Tweets that has the word "Jobs."

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

So now using the count vectorizer **[21]** we will transform the dataset into a bag of words, and then we will predict using the PKL file that has our best model to start predicting. All coding concepts used are mainly presented in the data mining domain **[8][9][10].** We do not need to fit our data because its already fitted using pickle, so only these steps are needed, then we transform the dataset into a new one with a column named Sentiment to check each tweet with its corresponding sentiment. As shown in **figure 100**, these are the tweets and their corresponding sentiments.

	Tweets	Sentiment
0	if we act now on the american jobs plan in years people will look back and say this was the moment that america won the future ‖url‖	1.0
1	‖target‖ imagine a future where we lead the world and tackle the threat of climate change with american jobs and ingenuity we can make	1.0
2	‖target‖ clean drinking water is infrastructure the american jobs plan will improve our water infrastructure by replacing of t	1.0
3	‖target‖ this week i announced the american jobs plan which will invest in our infrastructure and strengthen americas competitiveness	1.0
4	‖target‖ the american jobs plan will ensure every american can turn on the faucet or fountain and drink clean water ‖url‖	1.0
5	the american jobs plan is the largest american jobs investment since world war ii	1.0
6	under the american jobs plan of our nations lead pipes and service lines will be replacedso every child in america can turn on the faucet or fountain and drink clean water we ‖not‖ delay another minute	1.0
7	wall street didnt build this countrythe great american middle class did this time when we rebuild the middle class under the american jobs plan were bringing everyone along	0.0
8	delivering for the american people is what the american rescue plan was about its what the american jobs plan is about we can do this we have to do this we will do this	1.0
9	the american jobs plan is a once in a generation investment in america it will modernize miles of highways roads and main streets repair bridges desperately in need of upgrades replace of our nations lead pipes and service lines	1.0
10	millions of americans lost their jobs last year heres the truth we all do better when we all do well its time to build our economy from the bottom upand the middle outnot the top down	1.0
11	we ‖not‖ only have an economic imperative to act now we have a moral obligation in this pandemic in america we ‖not‖ let people go hungry we ‖not‖ let people get evicted we ‖not‖ watch nurses and educators lose their jobs we must act	0.0

Figure 103: Sentiment Detection of all Tweets related to Jobs posted by President Joe Biden

Sentiment 1 stands for positive and Sentiment 0 stands for negative. We can notice the tweet number 0 it says that if USA acts now towards the job plans, in the future, Americans will say that that was the moment where the US won and built its future successfully. Here Biden was confident and optimistic, and positive

and the sentiment was given as positive. The rest of the tweets as pretty much logical and should be with a true sentiment as well.

7.2- Images and Social media posts:

As used before with machine learning algorithms for sentiment analysis of pictures in **[92]** Testing was applied towards a picture of Dwayne the rock Johnson.

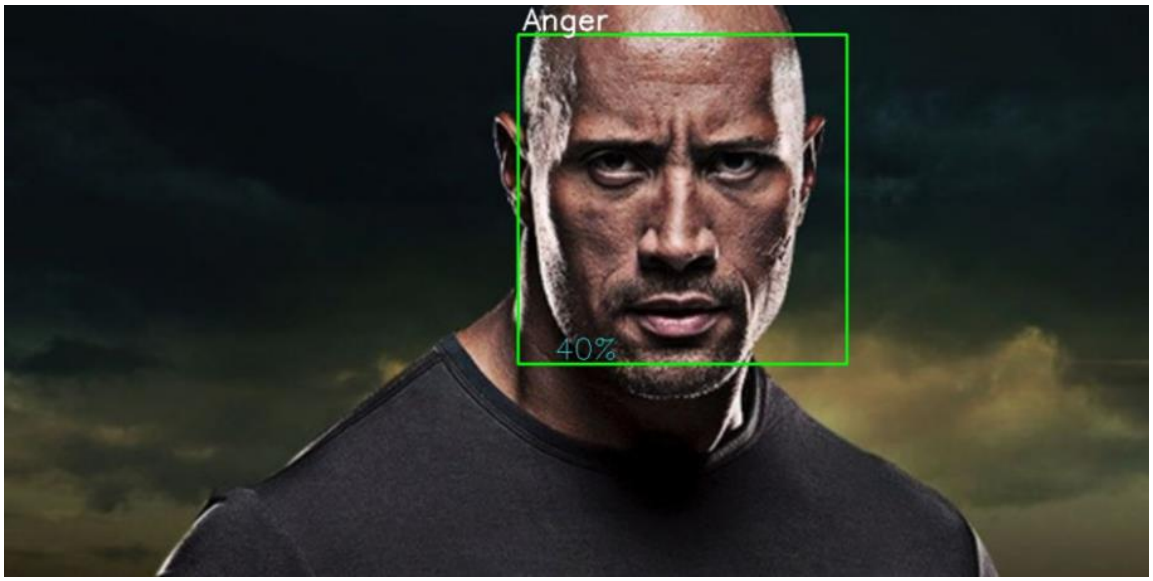


Figure 104: Applying ML to detect sentiment of a picture.

As you can see the sentiment analysis of the picture in **Figure 104** showed that he is 40% angry where the sentiment analysis procedure was used with 4 or 5 possible sentiments. Later in this report, we will be showing sentiment analysis using Deep learning and not machine learning, and we will test on the same picture, to see which approach is better to be used in this application.

7.3- Vital Signs:

The classification learner application trains the model to classify the data. With this program, you can use different classifiers to study supervised machine learning. You can check data, select features, define validation modes, train models, and evaluate results. Find the best type of classification model, including decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, ensemble classification and neural networks. Observations or examples) and known responses to the data (such as tags or classes). Use the data to train a model that generates predictions to act on new data. To use the model with new data or obtain program classification information, you can export the model to the workspace or generate MATLAB® code to reconstruct the trained model.

In Matlab, these are the steps that were followed to train the data:

1. On the "Applications" tab, in the "Machine Learning" group**,** click "ClassificationLearner".
2. Click "New Session" and select a workspace or file data. Specify the response variable and the variable used as the predictor variable. For information on classification issues, see the "Data Selection and Validation" section**.**
3. On the Classification Learner tab, under Model Type, click All Quick Training. This option trains all available preset models for your data set**,** and you can quickly customize them.

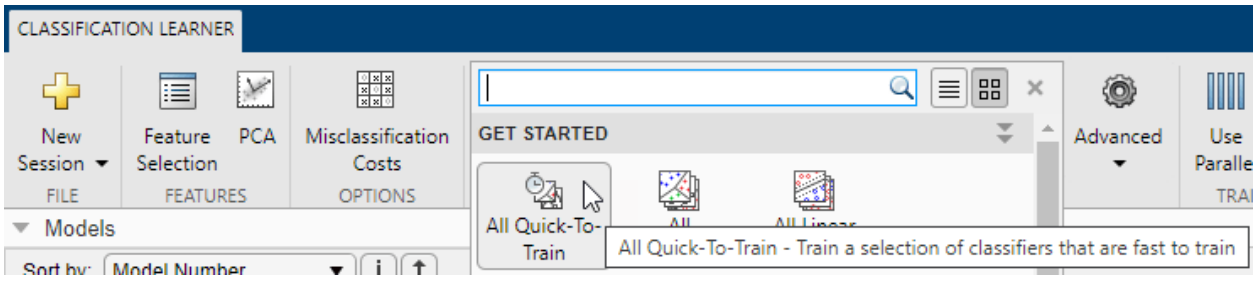


Figure 105:Quick train button

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

4. Click **Train**. 

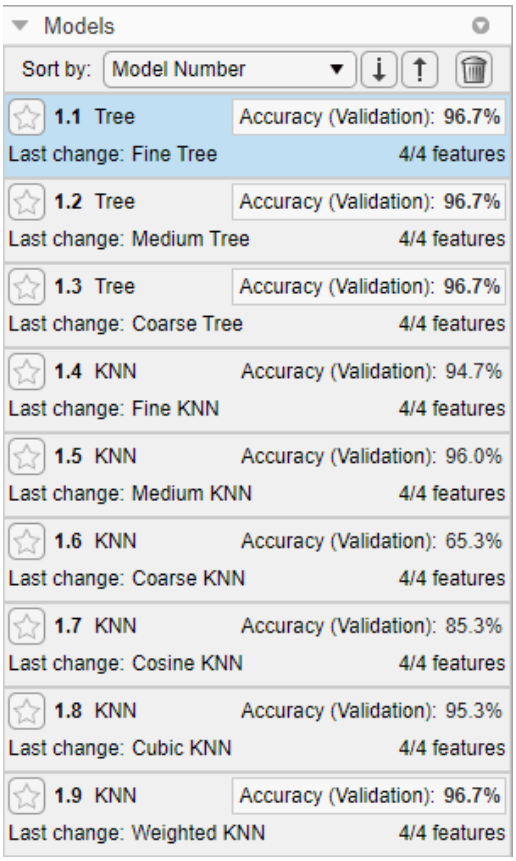


Figure 106:Classifier results

Complex tree classifier:

Decision trees or classification trees and regression trees can predict the answer to the data. To predict the answer, follow the solution from the root node (start node) to the end node in the tree. The leaf node contains the answer. The classification tree provides the answer. True or false in name. The regression tree provides numerical answers. The Statistics and Machine Learning Toolbox™ tree is binary. At each stage of

prediction, the value of the predictor variable is checked. For example, this is a simple classification tree:

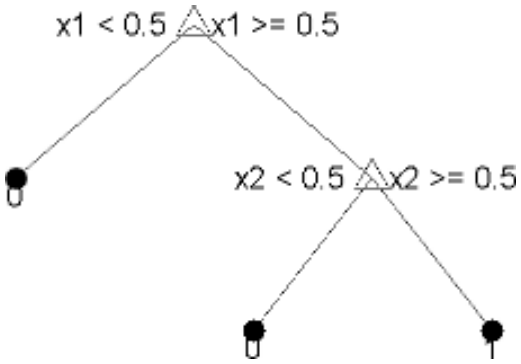


Figure 107: Decision tree

the tree predicts the classification based on two predictor variables x1 and x2. Start the prediction at the top node indicated by the triangle (Δ). The first solution is when x1 is less than 0. If so, follow the left branch and make sure the tree classifies the data as type 0. However, if x1 is greater than 0.5, move to the lower right triangle along the right branch. Here, the tree asks whether x2 is less than 0. If so, look at the tree along the left branch to classify the data as type 0. If not, follow the right branch to look at the tree to classify the data as type 1. **[119]**

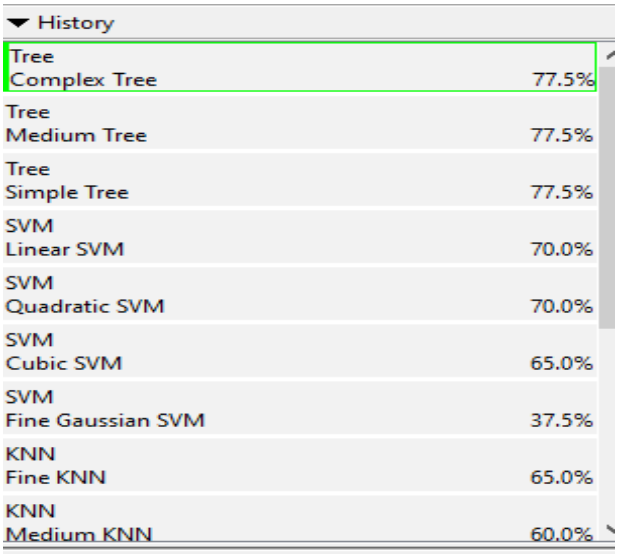


Figure 108: Classification results

The above steps were followed in our project. Data from the excel sheet, (10 subjects * 4 videos * 10 features = 400 cells) were imported to the matlab workspace. Classification learner was then accessed to classify these data. It is important to add the excel sheet as a table and add the response in the end of this table, in our case the response was (Angry, Happy, Sad, Neutral). After the addition of the excel sheet, response column was chosen and after many trials, the cross validation of 3 folds gave the best accuracies. The decision trees (complex, medium, simple trees) and the ensemble classifier of bagged trees gave the best accuracy which was of 77.5%.

Scatter plot: A scatter chart is a mathematical chart or a chart that uses Cartesian coordinates to

display the data values of two typical variables. Adjustment. The points are coded (color/shape/size) and other variables can be displayed. The data is displayed as a series of points, each of which has a variable value that determines the position on the horizontal axis and another value. A variable used to define the position on the vertical axis.**[121]**

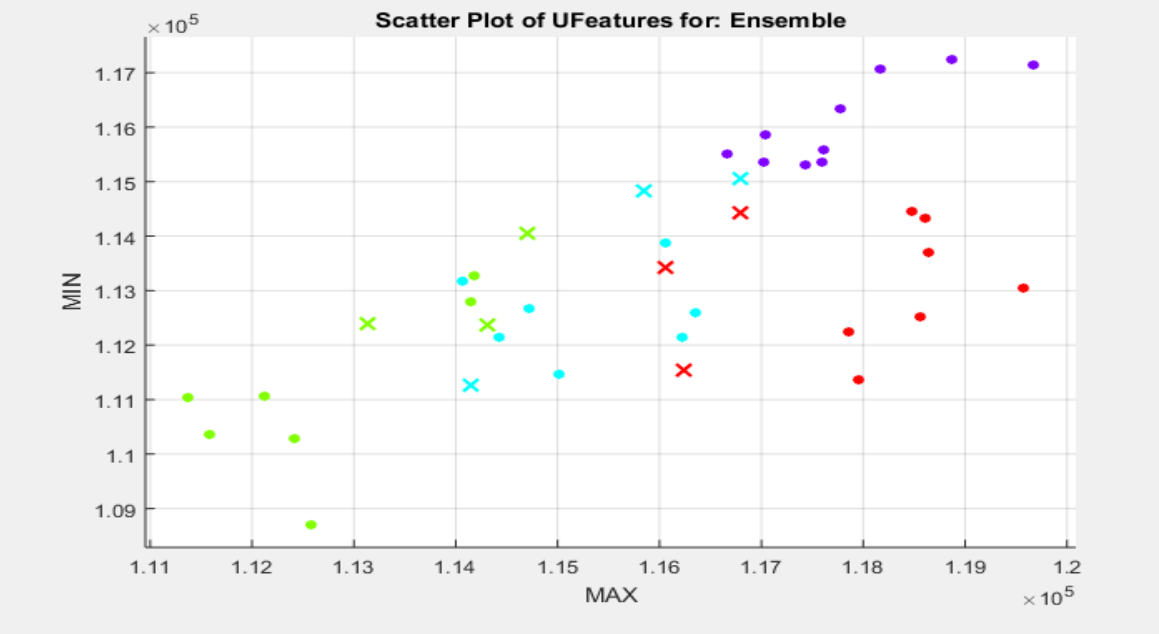


Figure 109: Scatter plot of features

In Figure 106, the plot is showing our results and the distribution of the points, each color represents a certain class and are classified due to the values that were calculate.

The confusion: matrix is an N x N matrix used to evaluate the performance of the classification model, where N is the number of target categories. The matrix compares the actual target value with the target value predicted by the machine learning model. **[122]**

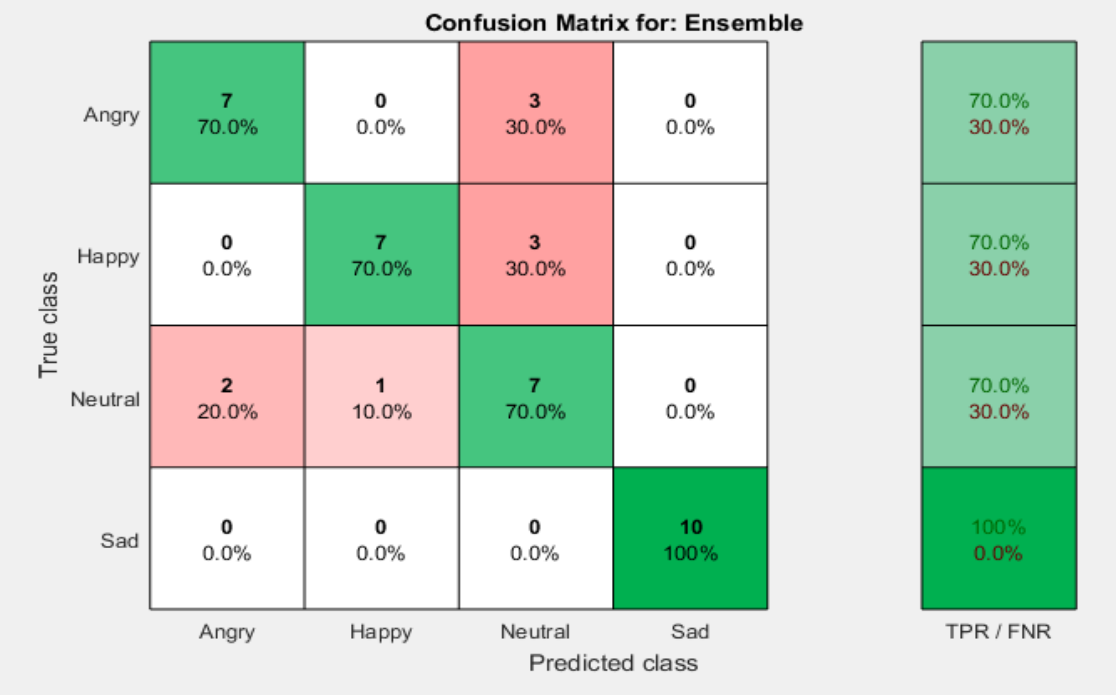


Figure 110: Confusion Matrix

In our results, this is the confusion matrix that was generated. A 70% accuracy between (Angry_Angry, Happy_Happy, Neutral_Neutral) was achieved and a 100% between sad_sad.

ROC curves were then obtained for each mood:

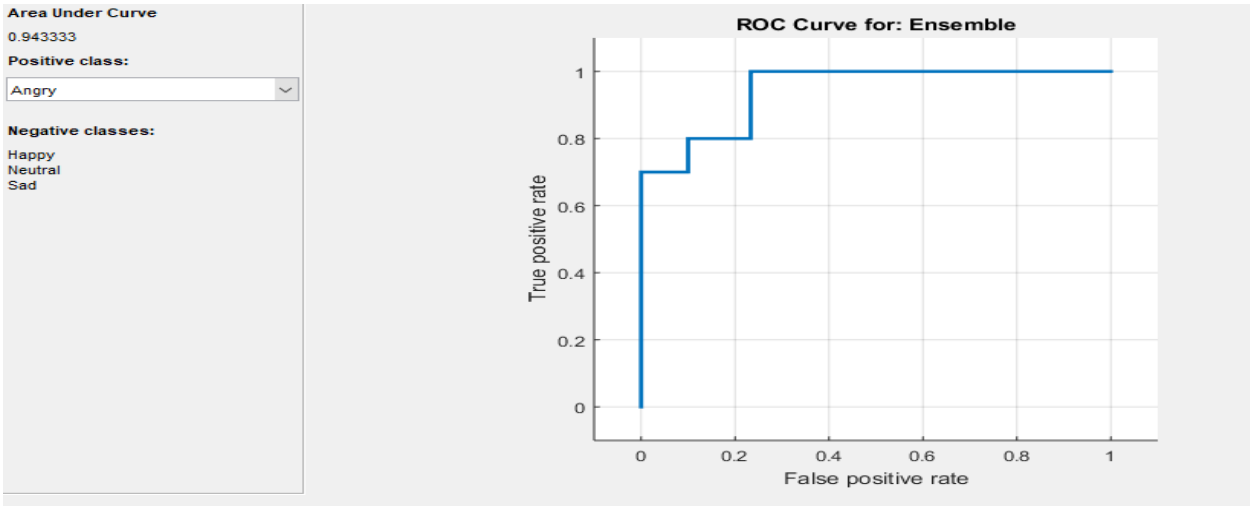


Figure 111: ROC for angry

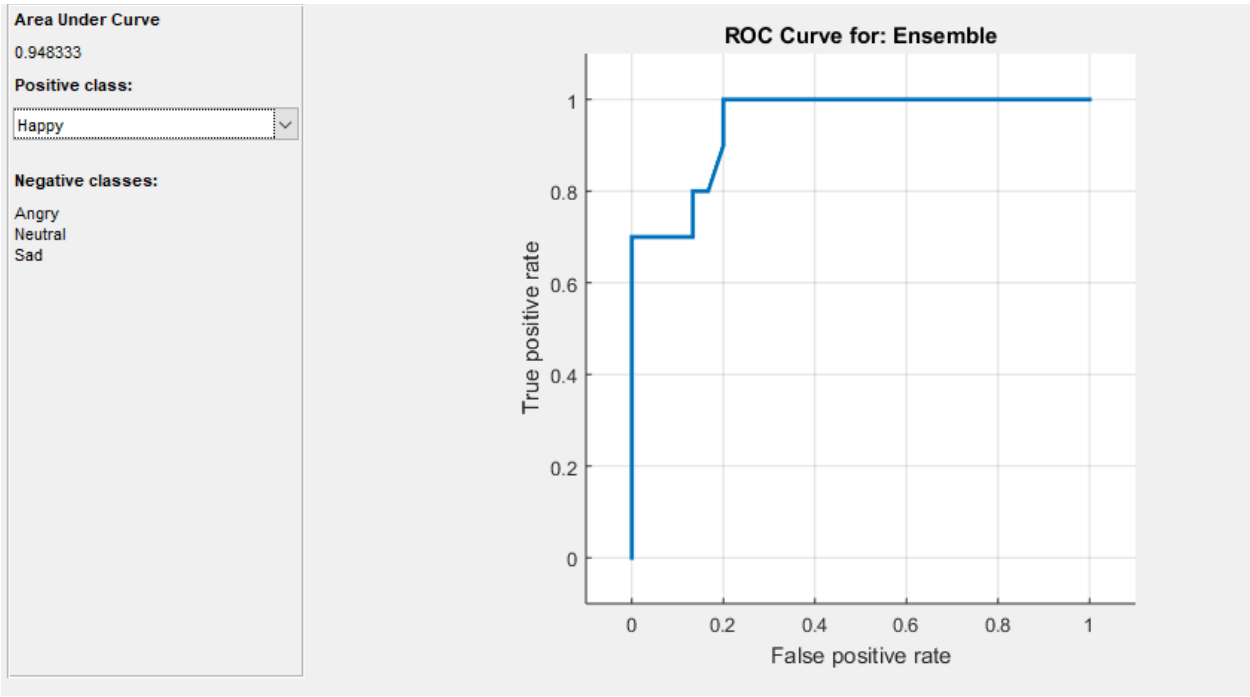


Figure 112: ROC for Happy

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

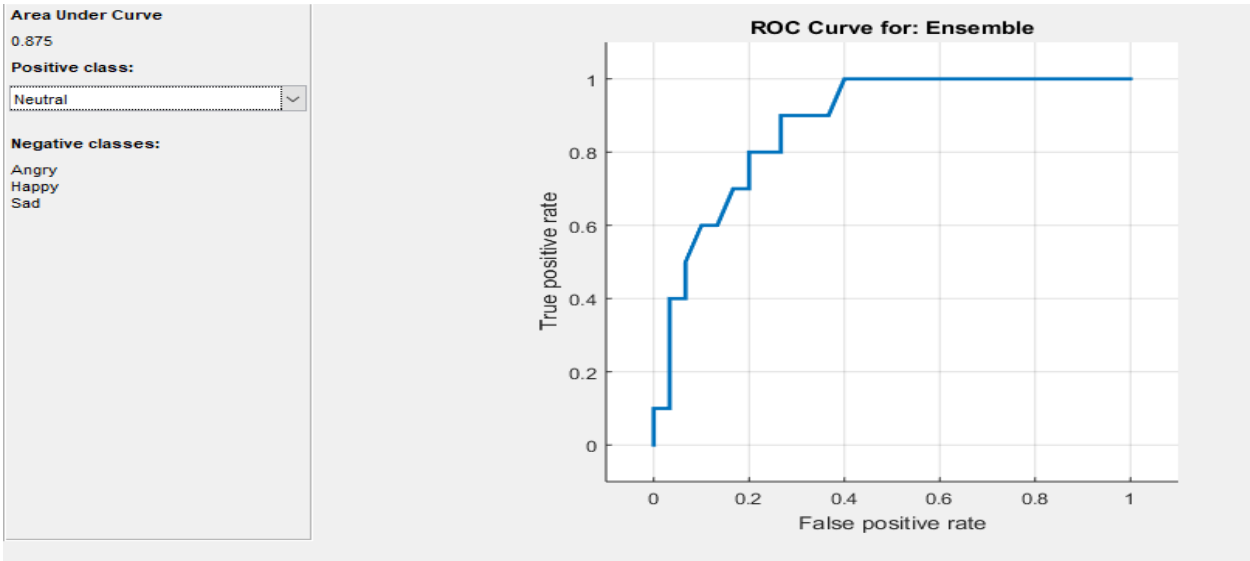


Figure 113: ROC for Neutral

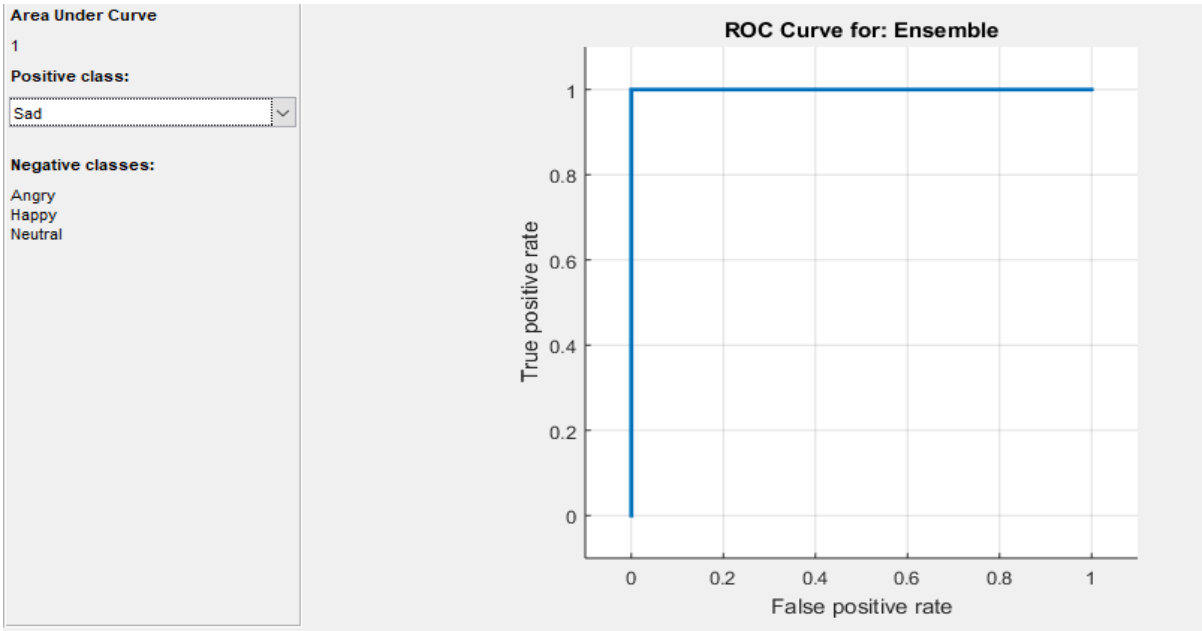


Figure 114: ROC for sad

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

The receiver response curve or ROC curve is a graphical diagram illustrating the diagnostic ability of the binary classifier system when its detection threshold changes. This method was originally developed for operators of military radar receivers, hence the name.

The ROC curve shows the balance between sensitivity (or TPR) and specificity (1-FPR). The classifier showing the curve closer to the upper left corner indicates better performance. The smaller the ROC space, the lower the accuracy of the test. **[123]**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

8- Deep Learning with Computer Vision:

[11] [12] [13] [14] [15]

8.1- Pictures and social media posts:

Since we used Machine Learning before, and we received a 40% angry on Dwayne Johnson picture, we will try to use Deep Learning with Computer Vision on the same picture and compare both approaches.

We start by importing CV2 library which is the computer vision library being used **[37][38].** We will also import the library DeepFace **[16][17][18].** This library is a library released in December 2020 by Facebook. It is basically a deep learning facial recognition system created by a research group in Facebook. It is considered one of the latest most trending libraries in the field of image processing, detection, verification, and analysis. This library can be installed in Jupyter notebook, and it identifies human faces in digital images. It employs a 9-layer neural network with over 120 million connection weights, in addition, this library was trained on four million images upload by Facebook 9users. It is stated that Deepface reaches approximately 97.35% plus or minus 0.25% on labeled faces in LFW dataset where human beings have 97.53%. So, DeepFace can sometimes be more successful than the human beings.

In addition to that, we imported Tkinter library **[35][36]** for a good GUI and flexibility in usage and in picking images and videos to analyze and detect the sentiment. We also imported matplotlib **[42][43]** library for visualizing the images and data.

We start by choosing the Dwayne picture from desktop to analyze it and detect the sentiment.

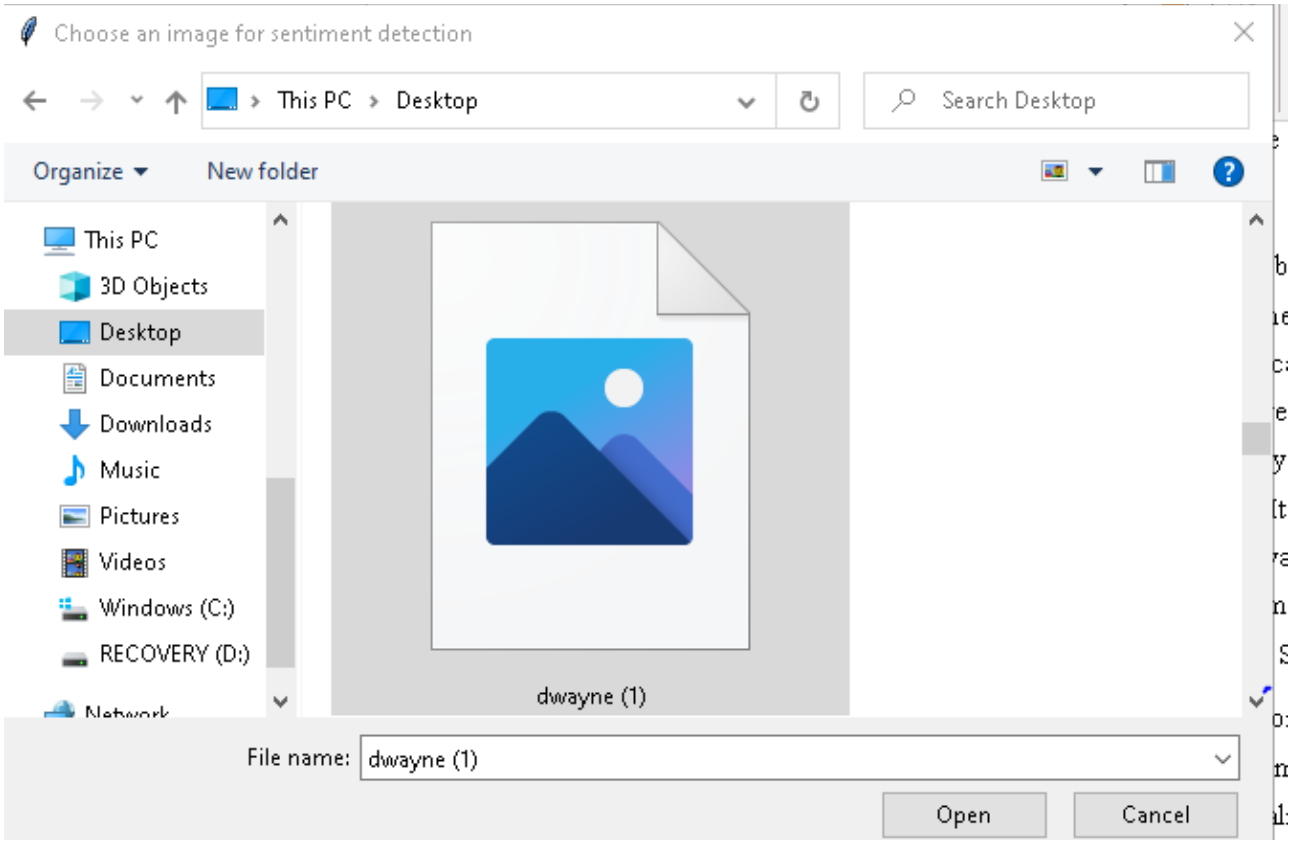


Figure 115: Choose the image for sentiment detection.

We notice below the picture is shown in blue color and its not very clear.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

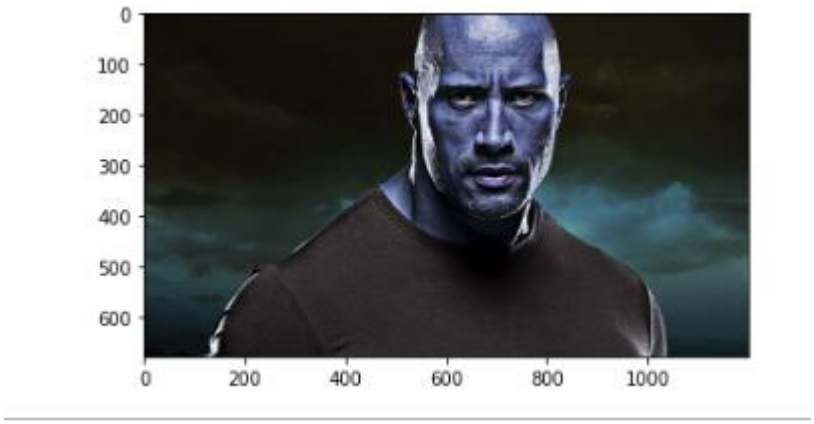


Figure 116: Retrieved picture for sentiment analysis.

So, we use computer vision to convert it from BGR to RGB format, and then we print it out using matplotlib as shown in **Figure 117.**

Figure 117: The original image color

After we have our image ready. We use one of DeepFace's famous functions which is **.detectface**, to detect the facial landmarks and the face expressions to be used in sentiment analysis with a better performance than performing the detection on the original image. You can notice the detected face in **figure 118.**

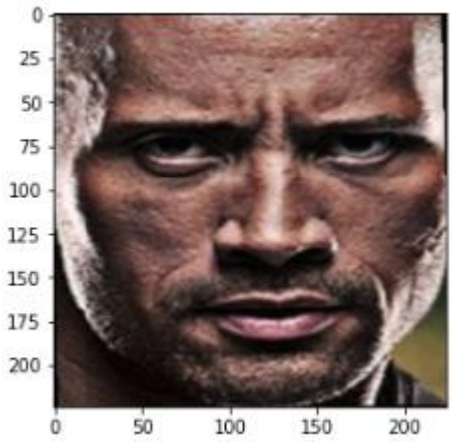


Figure 118: Detected face

Now after we detected the face, we will use another one of DeepFace functions which is **.analyze** so we start the sentiment analysis of the image. Now we can notice in **figure 119** the results.

```
Out[6]: {'emotion': {'angry': 81.87503110617584,
           'disgust': 0.0017335610333344955,
           'fear': 0.05900144812287081,
           'happy': 2.045568200681698,
           'sad': 3.0329700504523545,
           'surprise': 2.7239924057717566,
           'neutral': 10.261705384356105},
          'dominant_emotion': 'angry'}
```

Figure 119: Sentiment Detection results of the picture

As we notice before in the Machine Learning section, the sentiment analysis gave a 40% angry, whereas here it gave us 81.8% angry among 7 sentiments as well. The difference clearly shows that Deep Learning and especially DeepFace overcomes the Machine learning approach in sentiment analysis of images with no

doubt.

We will try our function on a picture of President Donald Trump.

As we notice below it gave us the dominant emotion as angry **Figure 120**, and that is false, so watch out when using the function to multiply the detected face image by 255, to get a suitable size for the detection phase.

```
Out[10]: {'emotion': {'angry': 81.91387272907821,
                      'disgust': 0.0015394314670844676,
                      'fear': 0.04952991722167525,
                      'happy': 1.9833784048702234,
                      'sad': 2.899286550368482,
                      'surprise': 2.6193090386022284,
                      'neutral': 10.533076759210019},
                     'dominant_emotion': 'angry'}
```

Figure 120: Sentiment Analysis of Donald Trump image before multiplying the detected face by 255.

Now after we multiply the detected face, we can notice the change in the results in **figure 121.**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

```
Out[11]: {'emotion': {'angry': 5.512802570070509e-05,
           'disgust': 1.659071770572257e-10,
           'fear': 0.2643402583566,
           'happy': 99.10081039657469,
           'sad': 0.012334904734634044,
           'surprise': 0.0007496292271249312,
           'neutral': 0.6217135306640231},
          'dominant_emotion': 'happy'}
```

Figure 121: Sentiment Analysis of Donald Trump image after multiplying the detected face by 255.

The deep face library with its function that is .analyze has a lot of features other than sentiment like race and age that can be performed without identifying the parameter of action into only emotion, so we do not type any parameter, more information can be provided in **[16][17][18]**. We can see the accuracy of this library as seen in **figure 122.**

```
Out[9]: {'emotion': {'angry': 5.512802570070509e-05,
          'disgust': 1.659071770572257e-10,
          'fear': 0.2643402583566,
          'happy': 99.10081039657469,
          'sad': 0.012334904734634044,
          'surprise': 0.0007496292271249312,
          'neutral': 0.6217135306640231},
         'dominant_emotion': 'happy',
         'age': 40,
         'gender': 'Man',
         'race': {'asian': 0.001521287685498664,
          'indian': 0.00021565440717154566,
          'black': 6.079908857287422e-06,
          'white': 99.0330218116071,
          'middle eastern': 0.13130851258289963,
          'latino hispanic': 0.8339309196389818},
         'dominant_race': 'white'}
```

Figure 122: Full feature Analysis of Donald Trump image

As you notice, the results are shocking and highly close to reality, this library is a huge blow up and advancement in the field of image processing, deep learning with facial recognition and sentiment analysis.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

8.2- Videos and interviews:

As we noticed earlier in the report, precisely in the **Data Preparation** heading, the videos, and interviews section. We clearly cropped the two parts of Joe Biden's interview video **[91]** into 2 separate videos, were we needed to analyze Joe Biden sentiments during these two parts, these two parts included the first question which is a result our first cropped video, the video showed that Biden was asked "Do you think trump would win?" and the second question was "Do you think raising the taxes is a good idea in this crisis?. Here in the Deep learning and computer vision we will be predicting and testing each video separately. We will be decomposing each video into frames, we take each second in the video as one frame, for example: if the video's duration is 10seconds then we will have 10 frames, each frame will be as a picture, each picture we will use Deep Face and CV2 libraries **[16][17][18][37][38]** in our procedure, we will be using deep face function to detect the face of each frame and using the other function which is the analyze to detect the sentiment and analyze the frame and we will be saving the dominant emotion of the picture in a array to be averaged later , at then end, we will be averaging all frames, and then we will have a sentiment analysis of the whole video, each video by its own, and then we will detect the dominant emotion among all the frames and we will have a clear result of the dominant sentiment of the whole interview question.

We will be using video file clip function of the library which is called moviepy.editor **[51]** and we will also use arrays.

We will start by giving the user ability to choose how many videos he/she wants to split into frames. Then the user will have permission to type the path he/she wants to save the frames, as well as the detected faces.
For example: C:\Users\USER\Desktop\BE\Videos\Detected_Faces\

This is the path of the detected faces for example. After that the user will choose the video, he/she wants to divide into frames using the Tkinter library that gives him access to his desktop and choose the video **[35][36]** after that, we will use the computer vision library cv2 to capture the video, and then we will use videofileclip that we imported to detect the duration of the video in seconds. After that we use the computer vision cap.get(cv2.CAP_PROP_FPS) to detect the fps of the chosen video.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

After that, the user will have to choose the name of the frames. Then a while loop will be implemented <=float(clip.duration) which is the duration of the video, we will be reading the video, each one second we will capture and save it to the directory provided previously by the user , then we will detect the face of the image and save it in the directory provided before by the user, and then we will analyze the picture using deep face, and the dominant sentiment will be saved into an array to be used later when averaging all the sentiments of all frames of a video. Then the next frame will be incremented, and same procedure will be done until all frames of the video are analyzed, and then the while loop will jump to the next video, until all videos are satisfied and done with saving detected faces, frames, and results of the analyzing.

You can notice the sentiments presented in both videos in the figure 120. We printed out the content of each array that we used previously when we loop through the videos and their frames and stores the dominant sentiments.

```
Video Number 0 :
['sad', 'fear', 'fear', 'sad', 'sad', 'fear', 'sad']

Video Number 1 :
['neutral', 'fear', 'fear', 'angry', 'angry', 'fear', 'sad', 'neutral']
```

Figure 123: Sentiments of each frames in each video

Video number 0 which is the first video (first question) we notice that we have 7 sentiments, thus the video had 7frames, so the duration was 7seconds, for each frame we have its sentiment. Whereas for video number 1 we have 8 sentiments, thus having 8frames and the duration of the video is 8seconds.

After that we averaged the unique sentiments in each array content, so we check the count and divide by the total number of elements and then multiplied by 100, if any sentiment has 0% its automatically removed the new array that we will be using to print the results. But, for the sentiment that has a percentage more than 0, we stored in a new array to be printed out later. You can see in **figure 124** the sentiment that was removed.

```
Video Number 0:


angry was removed because its doesnt represnt a sentiment of the frame
happy was removed because its doesnt represnt a sentiment of the frame
surprise was removed because its doesnt represnt a sentiment of the frame
neutral was removed because its doesnt represnt a sentiment of the frame
disgust was removed because its doesnt represnt a sentiment of the frame


Video Number 1:


happy was removed because its doesnt represnt a sentiment of the frame
surprise was removed because its doesnt represnt a sentiment of the frame
disgust was removed because its doesnt represnt a sentiment of the frame
```

Figure 124: The unused sentiments in each video

Finally, we loop through the videos and print the number of frames, the duration and the dominant sentiments percentage in each video as shown in **figure 125.**

```
Video Number 0 :


Number of frames: 7
Duration: 7seconds
Dominant Sentiments:
['sad 57.14285714285714 % ', 'fear 42.857142857142854 % ']


Video Number 1 :


Number of frames: 8
Duration: 8seconds
Dominant Sentiments:
['angry 25.0 % ', 'sad 12.5 % ', 'fear 37.5 % ', 'neutral 25.0 % ']
```

Figure 125: Sentiment Analysis results of Joe Biden's video

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

9- Conclusion:

In conclusion, we clearly showed the motivations behind this project, sources that any individual need to have some knowledge about to succeed in such project, and a clear definition of the project as well. In this project, we provided a complete sentiment analysis system that detects the sentiment of an individual based on his/her tweets posted on twitter application, we clearly showed how to prepare the data, preprocess using Natural language processing and data mining, and clean the data, training, folding, testing and all required steps to reach the testing phase. We clearly showed how to choose the best model for such system, we tried more than 20 types of models, and we chose one model to test it on any user that is active on twitter application. In addition to that, we clearly provided information about the latest libraries and technologies being used to detect the sentiment of an individual with a very high accuracy and F1 scores, based on social media posts and pictures, and we also compared two different approaches in this domain (Deep learning with computer vision) and (Machine learning with computer vision) and we chose the best one which was Deep learning. Also, we clearly provided sentiment analysis of an individual based on his/her vital signs and signal coming out of the individual's body. Finally, we showed. How to detect the sentiment of an individual during a certain interview of a video, in a very professional manner provided in this report with much more to be accomplished in the future work that we did not have time to do currently. This system is widely used in all domains mainly: Investigations in crimes and other violent actions, interviews with politicians and role models, HR interviews when hiring employees, universities, schools, Psychotherapist's clinics when dealing with patients, malls when trying to see the most liked product by guests, courts when having a meeting about a certain crime with the judge and many other domains like relationship meetings, hospitals, and many others.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

10- Future work:

We had a lot of more additional work in mind, but we could not accomplish them in the time being. For sentiment analysis of tweets, we know that there are some challenges in this domain, like: sarcasm detection, domain dependence, mentions to users that might be highly important for example: why did Joe Biden mention CNN in his tweet and not ABC for example, and thwarted expressions as well. In addition to that, analyzing profiles of people who like, retweet, comment on a specific tweet, and analyze what they comment as well. These challenges can be dealt with in the preprocessing phase, this can boost our accuracy and F1 score using more advanced preprocessing techniques. In addition to that, we all know that twitter application has a lot of international users, so another future work in this domain is to use the system on other languages like Arabic for example, so having the ability to do sentiment analysis of tweets that are in Arabic language as well. Moreover, we intend to perform the emotion detection according to age ranges, genders, and different nationalities. Finally, having the ability to do sentiment analysis on other social media platforms like Facebook, LinkedIn, and others. Concerning the analysis of social media posts and images, a future work in this domain was to collect a new dataset that shows fake expressions presented by individuals, for example an individual is sad, but he/she shows a facial expression and landmark that they are happy. Concerning the sentiment analysis of interviews and videos, a future work was to automatically decompose the interview video into cropped parts according to each question, for example: using voice recognition techniques to identify when did the interviewer start talking and when did he/she finished talking and take that as a video, and do the same for the person we are interviewing for example: Joe Biden.In addition to that, using speech processing techniques that will help the user, when trying to do sentiment analysis of different cropped videos, so the sentiment being taken will be compared to the words being spoken by the person who is being interviewed, and after some comparisons, a lot of work can be done in that approach. Concerning vital signs, a very large dataset shall be contacted (around 100 subjects) which adds more information for the system to recognize and hence, better accuracy. The more data there are (subjects and features), the better understanding the system can have to assess the sentiment of subjects. Another thing is to implement this high accuracy classification system in a robot that can interact with humans. By that we mean, the robot must recognize the emotion of the human and reacts to each emotion in a different way. A small example would be, if the robot sensed a "Sad" emotion, it should change the tone of its' voice to a

more caring and loving sound. Also add some of the lines that reflects happiness and try to remove the "Sad" emotion. If it recognizes an "Angry" sentiment, a reaction would be to play some calming, relaxing music, or some advice to not hurt anybody around.

Finally, combining all these parts of the sentiment analysis into one complete system implemented in a robot, mobile application, or a website as well for other users to benefit from it.

11- Standards

Temperature sensor Standards - ASME B40.9

Respiratory Rate Sensor (oximeter) standards: ISO 80601-2-61:2011

Pressure sensor standards – ISO 9001

Heart rate sensor standards – ISO/IEC

Joint -Matlab Standards: MISRA C:2004 Rules MISRA C:2012 Directives and Rules CERT C Rules and Recommendations ISO/IEC TS 17961 Rules MISRA C++:2008 Rules JSF C++ Rules AUTOSAR C++14 Rules CERT C++ Rules

- Camera Standards: Resolution Measurements

– ISO 12233 Noise Measurements

– ISO 15739 Sensitivity Measurements

– ISO 12232 OECF (tone reproduction) Measurements

– ISO 14524 Shading Measurements

– ISO 17957 Geometric Distortion Measurements

– ISO 17850 Chromatic Displacement Measurements

– ISO 19084 67 Image Flare Measurements

– ISO 18844 Shooting Time Lag Measurements

– ISO 15781 Low Light Measurements

– ISO 19093 Image Stabilization Measurements

– ISO 20954-1 Color Characterization Test Procedures

– ISO 17321-1 Camera Testing Guidelines

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

12- References:

[0] https://github.com/ahmadshibly12/Sentiment-Analysis-System

[1] https://jupyter-notebook.readthedocs.io/en/stable/

Last accessed: March 2021

[2] https://www.python.org/

Last accessed: April 2021

[3] https://en.wikipedia.org/wiki/Natural_language_processing

Last accessed: February 2021

[4] https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32

Last accessed: April 2021

[5] https://en.wikipedia.org/wiki/Machine_learning

Last accessed: April 2021

[6] https://www.ibm.com/cloud/learn/machine-learning

Last accessed: March 2021

[7] https://www.expert.ai/blog/machine-learning-definition/

Last accessed: March 2021

[8] https://en.wikipedia.org/wiki/Data_mining

Last accessed: April 2021

[9] https://www.talend.com/resources/what-is-data-mining/

Last accessed: April 2021

[10] https://www.springboard.com/blog/data-mining-python-tutorial/

Last accessed: April 2021

[11] https://en.wikipedia.org/wiki/Deep_learning

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Last accessed: April 2021

[12] https://www.mathworks.com/discovery/deep-learning.html

Last accessed: March 2021

[13] https://www.pyimagesearch.com/start-here/

Last accessed: April 2021

[14] https://www.sas.com/en_us/insights/analytics/computer-vision.html#:~:text=Computer%20vision%20is%20a%20field,to%20what%20they%20%E2%80%9Csee.%E2%80%9D

Last accessed: March 2021

[15] https://en.wikipedia.org/wiki/Computer_vision

Last accessed: April 2021

[16] https://pypi.org/project/deepface/

Last accessed: April 2021

[17] https://en.wikipedia.org/wiki/DeepFace

Last accessed: March 2021

[18] https://pypi.org/project/deepface/0.0.1/

Last accessed: March 2021

[19] https://scikit-learn.org/stable/

Last accessed: April 2021

[20] https://en.wikipedia.org/wiki/Scikit-learn

Last accessed: April 2021

[21] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

Last accessed: April 2021

[22] https://scikit-learn.org/stable/modules/tree.html

Last accessed: February 2021

[23] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

Last accessed: March 2021

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[24] https://scikit-learn.org/stable/modules/sgd.html

Last accessed: April 2021

[25] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Last accessed: April 2021

[26] https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

Last accessed: April 2021

[27] https://scikit-learn.org/stable/modules/svm.html

Last accessed: April 2021

[28] https://scikit-learn.org/stable/model_selection.html

Last accessed: April 2021

[29] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html

Last accessed: April 2021

[30] https://docs.tweepy.org/en/latest/

Last accessed: March 2021

[31] https://docs.python.org/3/library/re.html

Last accessed: April 2021

[32] https://en.wikipedia.org/wiki/Regular_expression

Last accessed: March 2021

[33] https://www.nltk.org/

Last accessed: March 2021

[34] https://en.wikipedia.org/wiki/Natural_Language_Toolkit

Last accessed: March 2021

[35] https://docs.python.org/3/library/tkinter.html

Last accessed: April 2021

[36] https://en.wikipedia.org/wiki/Tkinter

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Last accessed: February 2021

[37] https://docs.opencv.org/master/d6/d00/tutorial_py_root.html

Last accessed: February 2021

[38] https://en.wikipedia.org/wiki/OpenCV

Last accessed: April 2021

[39] https://pandas.pydata.org/

Last accessed: February 2021

[40] https://en.wikipedia.org/wiki/Pandas_(software)

Last accessed: April 2021

[41] https://pandas.pydata.org/docs/

Last accessed: April 2021

[42] https://matplotlib.org/stable/tutorials/introductory/pyplot.html

Last accessed: March 2021

[43] https://en.wikipedia.org/wiki/Matplotlib

Last accessed: April 2021

[44] https://docs.python.org/3/library/pickle.html

Last accessed: March 2021

[45] https://developer.twitter.com/en/docs/authentication/oauth-1-0a/obtaining-user-access-tokens

Last accessed: April 2021

[46] https://numpy.org/doc/

Last accessed: March 2021

[47] https://en.wikipedia.org/wiki/NumPy

Last accessed: April 2021

[48] https://en.wikipedia.org/wiki/FFmpeg

Last accessed: March 2021

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[49] https://ffmpeg.org/ffmpeg-all.html

Last accessed: April 2021

[50] https://kkroening.github.io/ffmpeg-python/index.html#ffmpeg.Stream.audio

Last accessed: March 2021

[51] https://zulko.github.io/moviepy/getting_started/getting_started.html

Last accessed: April 2021

[52] https://amueller.github.io/word_cloud/generated/wordcloud.WordCloud.html

Last accessed: March 2021

[53] https://docs.python.org/3/library/html.html

Last accessed: April 2021

[54] https://docs.python.org/3/library/collections.html

Last accessed: March 2021

[55] https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf

Last accessed: April 2021

[56] https://textblob.readthedocs.io/en/dev/#

Last accessed: April 2021

[57] https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/01.01-Help-And-Documentation.ipynb

Last accessed: April 2021

[58] https://www.tutorialspoint.com/google_colab/google_colab_quick_guide.html

Last accessed: April 2021

[59] https://docs.python.org/3/library/os.html

Last accessed: April 2021

[60] https://docs.python.org/3/library/time.html

Last accessed: April 2021

[61]https://docs.python.org/3/library/csv.html#:~:text=The%20csv%20module's%20reader%20and,the%20DictReader%20a

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

nd%20DictWriter%20classes.&text=The%20Python%20Enhancement%20Proposal%20which%20proposed%20this%20addition%20to%20Python.

Last accessed: April 2021

[62]https://www.cse.ust.hk/~rossiter/independent_studies_projects/twitter_emotion_analysis/twitter_emotion_analysis.pdf

Last accessed: April 2021

[63] Alexander Pak, Patrick Paroubek. 2010, Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Last accessed: April 2021

[64] Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant. Supervision, Last accessed: April 2021.

[65] Jin Bai, Jian-Yun Nie. Using Language Models for Text Classification. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau. Last accessed: April 2021

[66] https://github.com/marclamberti/TwitterEmotionAnalysis

Last accessed: April 2021

[67]https://nbviewer.jupyter.org/github/marclamberti/TwitterEmotionAnalysis/blob/master/TwitterSentimentAnalysis.ipynb

Last accessed: April 2021

[68] Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. Emotion Detection from Text and Speech: A Survey. Social Network Analysis and Mining, Springer, 8(1):28, 2018, Last accessed: April 2021.

[69] https://www.sciencedirect.com/science/article/abs/pii/S1877750318311037

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Last accessed: April 2021

[70] Isidoros Perikos, and Ioannis Hatzilygeroudis. Recognizing Emotions in Text using Ensemble of Classifiers. Engineering Applications of Artificial Intelligence, 51:191–201, 2016., Last accessed: April 2021

[71] Alexander Maedche, Stefan Morana, Silvia Schacht, Dirk Werth, and Julian Krumeich. Advanced User Assistance Systems. Business Information Systems Engineering, 58(5):367–370, 2016.
Last accessed: April 2021

[72] S. Lalitha, Sahruday Patnaik, T. H. Arvind, Vivek Madhusudhan, and Shikha Tripathi. Emotion Recognition through Speech Signal for Human-Computer Interaction. Electronic System Design (ISED), 2014 Fifth International Symposium on, IEEE, pages 217–218, 2014 Last accessed: April 2021

[73] Kashfia Sailunaz, Tansel Ozyer, Jon Rokne, and Reda Alhajj. Text-Based Analysis ¨ of Emotion by Considering Tweets. Machine Learning Techniques for Online Social Networks, Springer, pages 219–236, 2018. Last accessed: April 2021

[74] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Computing Surveys (CSUR), 50(2):25:1–25:33, 2017 Last accessed: April 2021.

[75] Nancy Semwal, Abhijeet Kumar, and Sakthivel Narayanan. Automatic Speech Emotion Detection System using Multi-Domain Acoustic Feature Selection and Classification Models. Identity, Security and Behavior Analysis (ISBA), 2017 IEEE International Conference on, IEEE, pages 1–6, 2017. Last accessed: April 2021

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[76] https://arxiv.org/ftp/arxiv/papers/1601/1601.06971.pdf

Last accessed: April 2021

[77] Liu, S., Li, F., Li, F., Cheng, X., &Shen, H.. Adaptive cotraining SVM for sentiment classification on tweets. In Proceedings of the 22nd ACMinternational conference on Conference on information & knowledgemanagement (pp. 2079-2088). ACM,2013. Last accessed: April 2021

[78] Meng, Xinfan, et al. "Cross-lingual mixture model forsentimentclassification." Proceedings of the 50th AnnualMeeting of theAssociation for Computational LinguisticsVolume 1,2012
Last accessed: April 2021

[79] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on SocialMediaData", IEEE 14th International Conference onMobile Data Management,Milan, Italy, June 3 - 6, 2013,pp 91-96, ISBN: 978-1-494673-6068-5, Last accessed: April 2021

http://doi.ieeecomputersociety.org/10.1109/MDM.2013.

[80]https://www.researchgate.net/publication/337498446_Emotion_Detection_from_Tweets_using_AIT-2018_Dataset\

Last accessed: April 2021

[81] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay, "Enhanced SenticNet with Affective Labels for ConceptBased Opinion Mining," IEEE Intelligent Systems, vol. 28, no. 2, pp.

31–38, 2013 Last accessed: April 2021

[82] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection

in text streams by analyzing Twitter data," International Journal of Data

Science and Analytics, vol. 7, no. 1, pp. 35–51, 2019. Last accessed: April 2021

[83] S. Yuan, H. Huang, and L. Wu, "Use of Word Clustering to Improve

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Emotion Recognition from Short Text," Journal of Computing Science and Engineering, vol. 10, no. 4, pp. 103–110, 2016. Last accessed: April 2021

[84] A. Seyeditabari, S. Levens, C. D. Maestas, S. Shaikh, J. I. Walsh, W. Zadrozny, C. Danis, and O. P. Thompson, "Cross Corpus Emotion Classification Using Survey Data," This paper was presented at AISB, 2017. Last accessed: April 2021

[85] M. Hasan and E. Rundensteiner and E. Agu, "Emotex: Detecting Emotions in Twitter Messages," 2014. Last accessed: April 2021

[86] https://web.cs.wpi.edu/~emmanuel/publications/PDFs/C30.pdf
Last accessed: April 2021

[87] Munmun De Choudhury, Scott Counts, and Michael Gamon, "Not all moods are created equal! exploring human emotional states in social media," in Sixth International AAAI Conference on Weblogs and SocialMedia (ICWSM'12), 2012. Last accessed: April 2021

[88] https://arxiv.org/pdf/1812.04510.pdf
Last accessed: April 2021

[89] M.H. Siddiqi, R. Ali, A.M Khan, Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans. Image Proc. 2015, 24, 13861398. [CrossRef] [PubMed] Last accessed: April 2021.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[90] https://github.com/misbah4064/facial_expressions
Last accessed: April 2021

[91] https://www.youtube.com/watch?v=kSAo_1mJg0g&t=159s
Last accessed: April 2021

[92]https://github.com/misbah4064/facial_expressions.git
Last accessed: April 2021

[93] https://www.mathworks.com/products/matlab.html Last accessed: April 2021.

[94] https://www.arduino.cc/en/software Last accessed: April 2021.

[95] https://code.visualstudio.com/ Last accessed: January 2021.

[96] https://github.com/ Last accessed: February 2021.

[97] https://www.mathworks.com/help/stats/classificationlearner-app.html Last accessed: January 2021.

[98] https://www.katranji.com/Item/ARDUINO-SENSOR-PULSE-OXIMETER-MAX30102 Last accessed: January 2021.

[99] https://www.katranji.com/Item/ARDUINO-UNO-REV3-CHINA Last accessed: January 2021.

[100]https://github.com/sparkfun/SparkFun_MAX3010x_Sensor_Library/blob/master/examples/Ex a

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

mple8_SPO2/Example8_SPO2.ino Last accessed: January 2021.

[101] https://www.arduino.cc/reference/en/libraries/sparkfun-max3010x-pulse-and-proximity-sensor-library/ Last accessed: January 2021.

[102] https://github.com/kontakt/MAX30100 Last accessed: January 2021.

[103] Kebe, M., Gadhafi, R., Mohammad, B., Sanduleanu, M., Saleh, H., & Al-Qutayri, M. (2020). Human Vital Signs Detection Methods and Potential Using Radars: A Review. *Sensors (Basel, Switzerland), 20*(5), 1454. Last accessed: January 2021.https://doi.org/10.3390/s20051454

[104] Dzedzickis, Kaklauskas, & Bucinskas. (2020). *Human Emotion Recognition: Review of Sensors and Methods. Sensors, 20(3), 592.* doi:10.3390/s20030592 Last accessed: January 2021.

[105] Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N.-K., & Lee, B. (2017). *Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine. Sensors, 17(10), 2435.* doi:10.3390/s17102435 Last accessed: January 2021.

[106] Shu, Yu, Chen, Hua, Li, Jin, & Xu. (2020). *Wearable Emotion Recognition Using Heart Rate Data from a Smart Bracelet. Sensors, 20(3), 718.* doi:10.3390/s20030718 Last accessed: January 2021.

[107] https://www.scitepress.org/Papers/2015/52411/52411.pdf Last accessed: April 2021.

[108] https://github.com/kontakt/MAX30100 Last accessed: January 2021.

[109] https://www.electroniclinic.com/max30100-pulse-oximeter-arduino-code-circuit-and-programming/ Last accessed: January 2021.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[110]https://www.google.com/search?q=maximum+amplitude+of+a+signal&rlz=1C1CHBD_enLB887LB888&hl=en&sxsrf=ALeKk01OmCkqUP188CXc9yJ4-GV7T0xb8g:1619214527344&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjNit6VrJXwAhUNnxQKHRRcCYkQ_AUoAXoECAEQAw&biw=1517&bih=730#imgrc=GVHP-IRafdwuLM Last accessed: January 2021.

[111]https://www.google.com/search?q=minimum+point+in+a+signal&rlz=1C1CHBD_enLB887LB888&hl=en&sxsrf=ALeKk03OgZOCpM3PFI55TusD2Du2pEVBXg:1619218334201&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiGnP6supXwAhVB7OAKHdKmCaIQ_AUoAXoECAEQAw&biw=1517&bih=730#imgrc=64k9xnw86yRAVM Last accessed: January 2021.

[112]https://www.google.com/search?q=median+in+a+signal&source=lmns&bih=730&biw=1517&rlz=1C1CHBD_enLB887LB888&hl=en&sa=X&ved=2ahUKEwjk7viAu5XwAhUJRhoKHfO_DHcQ_AUoAHoECAEQAA Last accessed: April 2021

[113]https://www.google.com/search?q=kurtosis+in+a+signal&source=lmns&bih=730&biw=1517&rlz=1C1CHBD_enLB887LB888&hl=en&sa=X&ved=2ahUKEwiJhrWYvJXwAhUMWxoKHRcVBScQ_AUoAHoECAEQAA Last accessed: January 2021.

[114] https://www.kdnuggets.com/2018/05/skewness-vs-kurtosis-robust-duo.html#:~:text=Skewness%20outputs%20values%20which%20are,(aka%20negatively%20skewed)%20signals. Last accessed: January 2021.

[115] https://www.raeng.org.uk/publications/other/8-rms#:~:text=The%20RMS%20value%20is%20the,a%20steady%206V%20DC%20supply. Last accessed: January 2021.

[116]https://www.dspguide.com/ch2/2.htm#:~:text=The%20average%20deviation%20of%20a,terms%20would%20average%20to%20zero.https://www.dspguide.com/ch2/2.htm#:~:text=The%20average%20deviation%20of%20a,terms%20would%20average%20to%20zero. Last accessed: January 2021.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

[117] https://en.wikipedia.org/wiki/Covariance Last accessed: January 2021.

[118] https://www.nlm.nih.gov/nichsr/stats_tutorial/section2/mod8_sd.html Last accessed: January 2021.

[119] https://www.mathworks.com/help/stats/decision-trees.html Last accessed: January 2021.

[120] https://www.mathworks.com/solutions/machine-learning.html Last accessed: January 2021.

[121]https://en.wikipedia.org/wiki/Scatter_plot#:~:text=A%20scatter%20plot%20(also%20called,for%20a%20set%20of%20data. Last accessed: January 2021.

[122] https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 Last accessed: January 2021.

[123] https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/ Last accessed: January 2021.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

13- Appendices:

13.1- Proposal Form

Year	2020-2021	**Date**	November 5, 2020

Project Title
Sentiment Mining

Project Synopsis
• The idea of this project is to detect someone's mood based on: social media pictures and images, tweets, interview videos and vital signs.

Project Category
• **Biomedical & Computer Communication Engineering**

Project Advisors
Dr. Rached Zantout Dr. Samir Berjawi

Number of Students
Two students

Required skills to complete the project
• The project needs good software practices and programming skills. • Knowledge in Natural Language Processing and Computer vision. • It needs knowledge to acquire vital signals from individuals. • Hardware skills to build circuits and run the sensors. • Python and MATLAB skills • Research capabilities. • Machine and Deep Learning Skills.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Expected Outcomes
• To have a preliminary design. • Finish the literature review. • Implement the system

Realistic Constraints
• Covid-19 and closure of university. • Economic crisis and inflation of the currency.

Applicable standards
- **pressure sensor standards:** ASTM Standards : D3951 Practice for Commercial Packaging ISO Standards : ISO 9001 Quality SystemModel for Quality Assurance in Design/Development, Production, Installation, and Servicing. (https://www.astm.org/Standards/F2070.htm) - **ECG (heart rate) standards**: ISO/IEC Joint (https://www.sciencedirect.com/science/article/abs/pii/S0022073619303942#:~:text=A%20new%20standard%20for%20electrocardiographic,including%20diagnostic%20equipment%2C%20monitoring%20equipment%2C) - **Temperature Sensor Standards:** ASME B40.9 (http://www.burnsengineering.com/local/uploads/content/files/Standards%20Defining%20Temp%20Sensors%20Notes.pdf)

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

- **Respiratory Rate Sensor (oximeter) standards:**

ISO 80601-2-61:2011 (https://www.iso.org/standard/51847.html)

- **Camera Standards:**

Resolution Measurements – ISO 12233

Noise Measurements – ISO 15739

Sensitivity Measurements – ISO 12232

OECF (tone reproduction) Measurements – ISO 14524

Shading Measurements – ISO 17957

Geometric Distortion Measurements – ISO 17850

Chromatic Displacement Measurements – ISO 19084

Image Flare Measurements – ISO 18844

Shooting Time Lag Measurements – ISO 15781

Low Light Measurements – ISO 19093

Image Stabilization Measurements – ISO 20954-1

Color Characterization Test Procedures – ISO 17321-1

Camera Testing Guidelines – ISO/TR 19247

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

13.2- Minutes of Meetings:

Minutes of Meeting (1)

February 10, 2021 | Started at 4:00PM | Location: Online via zoom.

Meeting called by Hasan Khamis.

Type of meeting: Informative

Facilitator: Dr. Zantout

Note taker: Hasan Khamis.

Attendees: Dr. Rached Zantout
Hasan Khamis
Ahmad Cheble
Dr. Samir Berjawi

Past Progress:

- The code for twitter sentiment analysis is being implemented and edited.
- Sentiment analysis of images is now being under study.
- System was virtually connected.
- Code was obtained from the Oxyllu library

Future Progress:

- Connect the system hardware.
- Prepare 4 separate videos that stimulate sentiments (Happy, Sad, Neutral, Angry)
- Contact subjects to take their heart rate.
- Implement the code for text classification with preprocessing.
- Implement the sentiment analysis of images.

Conclusion:

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Past and future progress was discussed. The meeting ended at 4:28PM.

Minutes of Meeting (2)
February 28, 2021 | *Started at 12:50PM | Location: Online via zoom.*

Meeting called by	Hasan Khamis	**Attendees**:	Hasan Khamis
Type of meeting	Informative		Dr. Samir Berjawi
Facilitator	Dr. Zantout		
Note taker	Hasan Khamis		

Past Progress

Hasan:

- System hardware was connected; however, the sensor wasn't turning on.

Progress Done

- System address was obtained as 0*57
- System address was changed to 0*15 using the VScode
- Library from SparkFun was recommended.

Conclusion:

After meeting for longer than an hour, Mr. Samir helped Hasan to learn on the VScode software and how to change the address of the sensor.

Minutes of Meeting (3)
March 3, 2021 | *Started at 4:00PM | Location: Online via zoom.*

Meeting called by	Hasan Khamis	Attendees:	Dr. Rached Zantout

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Type of meeting	Informative	Hasan Khamis
Facilitator	Dr. Zantout	Ahmad Cheble
Note taker	Hasan Khamis	Absences:Dr. Samir Berjawi_

Past Progress

Hasan:

- Implementing the MAX30100 connections on the breadboard.
- Downloaded VScode software and roamed freely in the downloaded libraries.
- Initialization was success but no detected vitals were recorded.
- Downloaded a library from SparkFun, Initialization was success and the sensor started recording data.

Ahmad:

- Implemented the code for text processing and now have a complete system to access twitter texts to detect the sentiment, testing will be done in the following weeks.
- Image processing is now being finalized to be able to give multiple sentiments with a percentage of each sentiment

Future Progress

Hasan:

- Collect data for 10 subjects.
- Collect the vital signs of the subjects while watching these videos.
- Determine the features and record them on an excel sheet.

Ahmad:

- Test and finalize the programs to detect sentiments from image and text.

Conclusion:

Past and future progress was discussed. The meeting ended at 4:30.

Minutes of Meeting (4)
March 24, 2021 | *Started at 4:02PM | Location: Online via zoom.*

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Meeting called by	Hasan Khamis	Attendees:	Dr. Rached Zantout
Type of meeting	Informative		Hasan Khamis
Facilitator	Dr. Zantout		Ahmad Cheble
Note taker	Hasan Khamis		Dr. Samir Berjawi

Past Progress

Hasan:

- Data was collected for 10 subjects for each video.
- Data was recorded and saved.
- Features (max, min, median, kurtosis,.) were calculated for each subject on each video and added to the excel sheet

Ahmad:

- Testing for text analysis is complete and the system can detect positive or negative sentiment.
- Image processing was also tested and based on the face features it gives the percentage for each sentiment.

Future Progress

Hasan:

- Add the excel sheet to matlab
- Access the classification learner tool and train the system using classifiers to get the best accuracy.

Ahmad:

- Document all the information and start writing the report.
- Prepare and test the sentiment analysis of interviews and videos.

Conclusion:

Past and future progress was discussed. The meeting ended at 4:24.

Minutes of Meeting (5)
April 4, 2021 | *Started at 4:04PM | Location: Online via zoom.*

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

Rafik Hariri University
جامعة رفيق الحريري

Higher Education as it should be.

Meeting called by	Ahmad Cheble	Attendees:	Dr. Rached Zantout
Type of meeting	Informative		Ahmad Cheble
Facilitator	Dr. Zantout		
Note taker	Ahmad Cheble		

Past Progress

Ahmad:

- Progress about tweets, social media posts and pictures were discussed and in addition to interviews and videos.
- Testing of the interview videos was presented as well

Future Progress

Ahmad:

- Document everything, results, methodology, and write them in the report.

Conclusion:

Past and future progress was discussed. The meeting ended at 4:40PM.

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

13.3- Schedule Form

SLP I Prelim. Phase I (1 Month)	3 Weeks	Group Formation		
		Subject Selection		**Launching of Litterature Review Phase**
	1 week	Advisor Meeting I		
SLP I Prelim. Phase II (3 Months)	11 Weeks	Research		
		Problematic		
		State of the Art		

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

			Survey Preparation	
			Advisor Meetings II, III	**SLP I_Fall 20 Term_Work Finishes**
SLP II Phase I (1 Month)		4 Weeks	Tasks Distribution	
			Required Materials	
			Launching Survey Filling	
			POs & PRs	
			Advisor Meetings IV, V	**Launching of SLP II's Design Phase**
SLP II Phase II (1.5 Month)		6 Weeks	Experimental Protocol	
			Methodology	
			Statiscal Analysis: Survey Data	
			Algorithm Development	
			Design Enhancement	
			Advisor Meetings V, VI	

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

					Implementation		
					Advisor Meetings VII		**Launching of SLP II's Results Phase**
	SLP II Phase III (3/4 Month)		3 Weeks		Testing		
					Adjustment		
					Advisor Meetings VIII		

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com

					Model Validation		**Launching of SLP II's Deliverable Phase**
	SLP II Phase IV (1/2 Month)		2 Weeks		Writing Report		
					Preparing Presentation		
					Advisor Meetings IX		
					Assessment		**SLP II_Spring 19 Term_Work Finishes**

Ahmad-shibly@hotmail.com

Hasankhamis10.5@gmail.com