

# USING NEURAL NETWORKS FOR FACIAL SENTIMENT ANALYSIS

By  
Aaron Ward

Supervisor(s): Stephen Sheridan

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
B.SC IN COMPUTING  
AT  
INSTITUTE OF TECHNOLOGY BLANCHARDSTOWN  
DUBLIN, IRELAND  
2018

## **Declaration**

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of **B.Sc in Computing** in the Institute of Technology Blanchardstown, is entirely my own work except where otherwise stated, and has not been submitted for assessment for an academic purpose at this or any other academic institution other than in partial fulfillment of the requirements of that stated above.

Dated: 2018

Author:

---

Aaron Ward

# **Abstract**

# Acknowledgements

# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Table of Contents</b>	<b>iv</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Abbreviations</b>	<b>x</b>
<b>1 Introduction and Background</b>	<b>1</b>
<b>2 Literature Review</b>	<b>3</b>
2.1 Analyzing and Detecting Employee’s Emotion for Amelioration of Organizations . . . . .	3
2.2 Deep Learning for Video Classication and Captioning . . . . .	4
2.2.1 Image-Based Video Classication using CNN’s and LSTM’s . . . . .	5
2.3 Subject independent facial expression recognition with robust face detection using a convolutional neural network . . . . .	6

2.4	Neuromarketing - The Art and Science of Marketing and Neurosciences Enabled by IoT Technologies . . . . .	7
2.5	Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order . . . . .	8
2.5.1	Facial Expression Recognition System . . . . .	9
2.5.2	Experiments . . . . .	10
2.6	Relevant Work and Critical Analysis . . . . .	11
2.7	Conclusion of Literature Review . . . . .	14
<b>3</b>	<b>Analysis</b>	<b>15</b>
3.1	The Machine Learning Model . . . . .	16
3.1.1	Algorithm . . . . .	16
3.1.2	Programming Language . . . . .	17
3.1.3	Machine Learning Library . . . . .	18
3.2	Data Preparation . . . . .	19
3.2.1	Public Datasets . . . . .	19
3.2.2	Creating a Dataset . . . . .	20
3.3	Training . . . . .	20
3.3.1	Training Locally . . . . .	20
3.3.2	Cloud Training . . . . .	20
3.4	Hosting/Deployment . . . . .	21
3.5	Concluding Objective . . . . .	21
<b>4</b>	<b>System Design</b>	<b>22</b>
4.1	Machine Learning Model . . . . .	22

4.2	Data Preparation . . . . .	23
4.3	Training . . . . .	23
4.4	Deployment . . . . .	23
4.5	Concluding Remarks of System Architecture Design . . . . .	24
<b>5</b>	<b>Implementation</b>	<b>26</b>
5.1	Implementing the Machine Learning Model . . . . .	26
5.2	Deploying the Trained Model . . . . .	26
5.3	Implementing the Node.JS Application . . . . .	26
5.3.1	Database Integration . . . . .	26
5.4	Result Weighting Algorithm . . . . .	26
5.5	Integrating Third Party Services . . . . .	26
5.6	Conclusion . . . . .	26
<b>6</b>	<b>Testing and Evaluation</b>	<b>27</b>
6.1	Testing Model with Testing Data . . . . .	27
6.2	API Testing . . . . .	27
6.3	Testing Model with New Data . . . . .	27
<b>7</b>	<b>Results and Discussion</b>	<b>28</b>
7.1	Results . . . . .	28
7.1.1	Model Accuracy . . . . .	28
7.1.2	Usability . . . . .	28
7.2	Discussion . . . . .	28

<b>8 Conclusion and Further Work</b>	<b>29</b>
8.1 Conclusion . . . . .	29
8.1.1 Further Work . . . . .	29
<b>Bibliography</b>	<b>30</b>
<b>Appendices</b>	<b>33</b>
A Code Snippets . . . . .	33
B Screen shots . . . . .	33
C Neural Network Diagram . . . . .	33



# List of Tables

2.1	Test cases by Lopes et al. . . . .	11
-----	------------------------------------	----

# List of Figures

3.1	Graph of Machine Learning or Data Science Languages for 2016 by Verma (2017) . . . . .	18
4.1	System architecture of the prototype . . . . .	24

# Abbreviations

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DBN	Deep Belief Networks
LSTM	Long Short-Term Memory
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
ML	Machine Learning
NLP (1)	Neuro-Linguistic Programming
NLP (2)	Natural Language Processing

# Chapter 1

## Introduction and Background

The field of artificial intelligence and machine learning has become exponentially prominent in our daily lives. From business to social media, machine learning algorithms are being used to change the definition of efficiency and user experience. For example, Artificial intelligence algorithms are utilised by many large companies optimize the experience with image and voice recognition and photo searching (Deshpande, 2016). Also, Motor companies such as Tesla Motors use computer vision for their self-driving cars, which is a form of artificial intelligence.

The researcher and founded of convolutional neural networks (CNN's), Yann LeCun, became the director of Facebook's Artificial Intelligence department in 2013, and it is said believed that Facebook uses CNN's for it's facial recognition, user classification and tagging features (Deshpande, 2016). CNN's, in conjunction with recurrent nets, are also used for Facebooks DeepText feature. DeepText is a deep learning text-understanding engine used to comprehend and classify human generated textual content in over 20 languages (Abdulkader et al., 2017). In recent year, neural networks have been used in the field of medicine to better predict diagnoses and detection of cancerous tumours. For example, CNN's have been used by researchers for brain tumor segmentation (Havaei et al., 2015). The use of CNN's proved to be an appropriate method for tumour segmentation as the results can be given from a range of 25 seconds to 3 minutes (Havaei et al., 2015). Artificial neural networks have been used by radiologists for Computer-Aided detections systems (CADe) and Computer-aided Diagnosis

systems (CADx) to improve the accuracy of diagnoses, early detections and to minimize the time spent on evaluation by doctors (Firmino et al., 2014).

Many businesses depend on artificial neural networks for their business model as they can be applied to many industries and disciplines. According to Bhargava and Gupta (2017) artificial neural networks are used in a range of business applications such as forecasting of sales, classification of spending patterns, market targeting, risk analysis and bankruptcy prediction, to name a few.

# Chapter 2

## Literature Review

### **2.1 Analyzing and Detecting Employee's Emotion for Amelioration of Organizations**

Subhashini and Niveditha make the opening statement that emotions usually do not any place in a work environment in current society. Although the expression of feelings is suppressed in places of work, they suggest that emotions can affect five major areas in competitive advantage. The five given aspects of competitive advantage are as follows: Intellectual Capital, Customer Service, Organizational Reactivity, Production, Employee appeal and retentivity (Subhashini and Niveditha, 2015). In order to counter this apprehensiveness to expression of emotions in the work place, Subhashini and Niveditha suggest the concept of a facial emotion tracking system that will map the facial expressions of an employees face as they enter the organization. Two related works are also given; Emotion Detections Based on Text and Emotion Recognition Based on Brain-Computer Interface Systems.

The system architecture given by Subhashini and Niveditha briefly describes the program. Most employees entering a building to an organization must swipe a card to clock into the work hours. They suggest that they have designed a new system that removes the need for card swiping but also implements emotion detection. The employee must look into a camera that will prove their presence in the vicinity but will also perform some emotion detection. The system was implemented in the C Sharp programming language and uses

skin tone segmentation to detect. The binary image is then converted to an RGB image and an inspection of every individual pixel is performed. If the RGB value is greater than 110, then the pixel colour is refactored to be a white pixel, otherwise it becomes a black pixel. This is done to make it easy to detect facial features in the video stream (Subhashini and Niveditha, 2015). Once detected, the image around the face is cropped. They then apply a Bezier Curve to the regions around the lips and eyes of the person being analyzed. The results of the person's identity and emotional status are then stored within a database. They conclude their paper by explaining that this system can be used by management to gain an understanding of their employee's sentimental state.

## 2.2 Deep Learning for Video Classification and Captioning

Wu et al.'s paper provides an in-depth analysis on the methods for video classification and video captioning in terms of deep learning. They claim that because of the exponential growth in internet bandwidth and computing power, video communications are becoming more and more prevalent, therefore paving the way for new video understanding applications (Wu et al., 2016). They make reference to current implementations to prove the growth of interest in the field of computer vision and video analysis, notably the ImageNet challenge.

Move over, they go on to give brief description of the two "deep learning modules" that have been used for visual analysis: Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). It is explained that LeCun et al. that developed LeNet-5 made a break through when they developed a CNN using the Back-Propagation Algorithm. But it is noted that this is limited in performance when the complexity of the tasks is increased. Deep belief networks were developed to train networks in a unsupervised manner in order to counter this problem (Wu et al., 2016). AlexNet, a CNN proposed by Krizhevsky et al. in 2012, introduced two ways to increase the performance of CNN's using ReLU (Rectified Linear Units) and Dropout to decrease overfitting (Wu et al., 2016). Secondly, RNN's are

brought forward. Wu et al. explain the difference between CNN's and RNN's, stating that CNN's are all feed forward networks that do not use cycling, which can prove be disadvantageous when working with sequence labeling. Two issues can occur with RNN's: Vanishing Gradient and Exploding Gradients as short term memory is used when cycling through the network. The solution given to this is an RNN variant called Long Short-Term Memory (LSTM).

### **2.2.1 Image-Based Video Classification using CNN's and LSTM's**

They state that Karparthy et al. researched the common architectures for learning spatial-temporal clues in large video datasets. It appeared that models using single frames as input achieve similar results as models using stacks of frames. From this, Simonyan and Zisserman proposed the idea of the Two-Stream Approach, because of the cost effectiveness and time consumption that come with training 3D CNN's. This Two-Stream approach involves training the CNN on single and stack frames concurrently. Both outputs are put through a score fusion. The result is the weighted sum of both scores (Wu et al., 2016).

Although Two-Stream is a good approach, it is not sufficient as it is not capable of dealing with long video clips. Therefore, LSTM's are utilized as they do not suffer from the problem of vanishing gradients. It has been found that CNN's and LSTM's compliment each other when working in conjunction with each other (Wu et al., 2016). They conclude their paper with a summary of the written topics about, regarding the growth for the need for video understanding applications, the used of CNN's and RNN's, in addition to the variants of these deep learning modules.



## **2.3 Subject independent facial expression recognition with robust face detection using a convolutional neural network**

As stated by Matsugu et al., difficulties may arise with facial recognition. In terms of a face being in a smiling-like state, could have different implications. As well as this, a facial recognition system should be able to work with a wide range of variability of faces. They then give some examples of past implementations such as facial recognition with rigid head movement by Black and Yackoob in 1995, and speak about how this does not meet the requirements of dealing with wide variance. A rule based system is proposed (Matsugu et al., 2003). With their model, layer trains on a module-by-module (module being the nose, eyes, mouth etc) basis. Meaning each layer trains on a certain facial feature. Each of the neurons perform an averaging of some local receptive fields then they use a skin tone detector to detect each module on the face (Matsugu et al., 2003). For training, Layer one and layer two are trained for 8 modules using back propagation. Layer three and four train on more complex feature detectors such as the mouth and eyes. The output is then sent to the rule based algorithm for handling variability and robustness.

The rule based algorithm takes the output of the CNN and measures the distance between the features. From these calculation, the rules are applied to determine if the person is in a laughing or smiling state. The rules are summarised as follows (Matsugu et al., 2003):

- The distance between eyes and lip get shorter.
- The horizontal length of lip gets longer.
- The eyes wrinkle.
- The gradient of lip from the end point to the end point increases.
- Detection of teeth increases.
- The edges (wrinkles) of cheek increase.

In conclusion, they received a 97.6 percent accuracy for 10 test subjects with 5600 images. They assert that their model is significantly more efficient as they only require one CNN due to their rule based algorithm, in contrast to Fasels implementation that uses two CNN's working in synergy with one another.

## **2.4 Neuromarketing - The Art and Science of Marketing and Neurosciences Enabled by IoT Technologies**

A paper by Arthmann and Li describes the growing field of Neuromarketing with an opening statement: Advertisers recognise that there is a relationship between stimulating the emotions of a customer and influencing their actions. Online shopping has drastically affected the store sales, and it is also explained that more than 3500 stores have shut down due to bankruptcy in 2017 (Arthmann and Li, 2017). Their answer to this change in buying is Neuro Linguistic Programming. Neuro Linguistic Programming (NLP), not to be confused with natural language processing, is a form of observing the verbal and non-verbal communication of humans. Eye accessing cues (eye movements) are said to be linked to certain emotions or thoughts. Neuromarketing incorporates NLP and IoT devices to understand the consumer sentiment more extensively. Neuromarketing aims to remove marketing biases by utilizing the consumers subconscious. One example given by Arthmann and Li is the notion of facial coding and motion tracking, which is put in place to determine why consumer make certain decisions. Furthermore, they make their belief clear of retailers benefiting from this when it is put forward that artificial intelligence and machine learning will evolve, giving better results of consumer preference shifts. Additionally, these neuromarketing systems can replace thermoimaging people counting devices that clock people walking into stores that may not be eligible customers (children). Further examples are provided for these technologies. The describe a scenario when a customer is given an image of a product and a price in front of a webcam, and by performing facial coding and sentiment analysis, we may get a better sense of what the consumer is feeling. Arthmann and Li conclude their paper by

declaring the future potential of these systems using when integrated with "always on" IoT technologies.

## **2.5 Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order**

Lopes et al. open their paper by explaining the definition of facial expression. They describe it as facial changes that occur with someone due to emotional state or social communication. In terms of facial expression recognition software, a lot of systems give misleading accuracy results due to the overlapping of training and test data. They explain that some problems may occur. For example: dealing with ethnicity and variability of faces. Their response to this is training with one data set and testing with another to provide more accurate results (Lopes et al., 2017). They make reference to Liu et al's work on facial recognition by describing the three stages of training: Feature learning, which is responsible for extraction of all facial features. Secondly, Feature Selection, that selects the best features to learn. And lastly, the classifier, that each expression has one specifically allocated to. Lopes et al. then move on to explain convolutional neural networks (CNN) on a high level. Firstly, the CNN is comprised of a convolution layer, that if given a kernel size. This kernel shifts along the image given to generate a map. This is followed by sub-sampling. Sub-sampling is used to reduce the map size to increase the accuracy in variance. Lastly, the fully-connected layer is introduced. This is a neural networks that has fully-connected neurons to it's previous layer (Lopes et al., 2017) and it is stated that CNN's using supervised learning use gradient descent. They explain that of all the facial recognition methods, CNN's prove to be very advantageous because they can use raw image data input for an accurate prediction.

A related work section is then brought forward to explain similar models that have been developed. Much progress has been made in the field of neural networks have come about in recent years. This is due to advances in GPU technologies and computing power Lopes et al.

(2017). One example of relevant work they provide is the work done by Song et al., in which a CNN was developed for a mobile phone application for facial expression recognition. This CNN used image augmentation techniques, due to the lack of public data for their network, to prevent over-fitting. This increases the amount of data available for training the network (Lopes et al., 2017). Song et al. received an accuracy of 99.2 percent using the CK+ dataset. A CNN with 15 layers was developed by Burkert et al that achieved similar result, of 99.6 percent. They point out that although this network achieved high results, they may prove to be misleading as it is not made clear that their training and test datasets were different (Lopes et al., 2017). They then go on to compare the related works by expressing the flaws of using over-lapping data and the lack of emotion expressions classified.

### **2.5.1 Facial Expression Recognition System**

Following the introduction and related work, Lopes et al. provide a prerequisite understanding for their model. Their model has two stages: the training stage and the testing stage. The training stage consists of a few preliminary steps. Firstly, new images are made from existing ones in the dataset to increase the trainable data. This is done using a method proposed by Simard et al. called "synthetic sample generation", which involves rotating and skewing the existing images. For every photograph that exist, an additional 70 are made, adding noise to the data. This synthetic data is only used in the training stage and advantageous as it allows the model to handle variance in an image (Lopes et al., 2017). Secondly, to address the problem in alignment with facial features, the notion of rotation correction is introduced that aligned the images using the eyes as the horizontal axis. Image cropping is then used to reduce the amount of background noise as it is said to decrease the accuracy and overall performance of the CNN. This is done by detecting only features that are valid of expression classification and cropping the image around them, excluding the neck, ears and background from the image. Down sampling is then applied for reducing the size of the image, making it 32 x 32 pixels in size. Brightness and contrast can cause problems with images, therefore intensity normalisation is applied to lower these aspects of the image (Lopes et al., 2017).

As for the testing stage, the same methods are used as the training stage. The CNN outputs the predicted emotional expression with the following number ID's:

- 0** - Angry
- 1** - Disgust
- 2** - Fear
- 3** - Happy
- 4** - Sad
- 5** - Surprise

Their model is comprised of 2 convolution layers, 2 sub-sampling layers and 1 fully-connected layer. The sub-sampling layers use max pooling with a 2x2 kernel. This halves the size of the image. The fully-connected layer consists of 256 neurons and provides 6 prediction outputs. Stochastic gradient descent is applied for back propagation and the Loss is calculated using soft-max. They also use the rectified linear unit (ReLU) for their activation function.

### **2.5.2 Experiments**

Lopes et al.'s experiments include a number of different test cases. These test cases utilize the CK+ database that has images of a hundred students ages from 18 to 30. An average accuracy is given for each of the test cases. See table 2.1 on page 11 for results.

Test Case Averages	
Test Case	Average
no pre-processing	53.57%
just cropping	71.67%
just rotation correction	61.55%
cropping and rotation	87.86%
intensity normalisation	57%
both normalisations	86.67%
spatial normalisation and synthetic data	87.1%
both normalisations and synthetic data	89.76%

Table 2.1: Test cases by Lopes et al.

In summary, it can be seen from the table above that by applying the synthetic and normalized data to the training dataset, a higher accuracy can be achieved when testing.

## 2.6 Relevant Work and Critical Analysis

Subhashini and Niveditha, 2015's work on analyzing and detecting employee's emotions for organizations ties in very well with the proposed project as it involves using sentiment analysis on subject to gain an underlining understanding of their emotions that may or may not be expressed verbally. Their approach to skin tone segmentation for detecting a humans face may prove very beneficial to the proposed project. The main research goal was to achieve employee identification in conjunction with facial emotional analysis and they were successful in execution, They used a Bezier Curve on the subjects lips and eye to detect the emotion expressed by analysing the gradient of the curve. However, Although this paper proved that the project was a success, there are some inconsistencies. For example: in the related work, they do not explain how the work is related and only give titles. Secondly, the results show no code snippets or pseudocode to explain the implementation of the system. Only screen shots of the user interface are provided. Furthermore, there are more absences of proof. The paper is lacking statistics and graphs to display the accuracy or progress of the systems performance and there are some bold statements used that are not back up by citations like when it is said that "emotions were considered a forbidden topic in the working

place”. Despite these weak areas in the paper, a good aspect of this system is the use of real life subjects used in testing.

The review of Deep Learning for Video Classification and Captioning by Wu et al. provides an in depth look into the aspects of different neural networks and what are they strong and weak points. The motivation for their research is driven by their claim that video communications is growing and that their needs to be better applications for video understanding. The relevancy of this paper provides the concept of the ”Two-stream” architecture. Although it is not planned to develop two convolutional neural networks (CNN), is it sought after to develop a score fusion algorithm, similar to the one mentioned in this paper. A CNN will be developed and the application will utilize a tone analyser for voice sentiment analysis and the two scores will be combined by such an algorithm to provide the weighted sum of the two scores. As this paper is very in depths and draws good comparisons between the different techniques that can be used for video classification. Despite the quality of this paper, some aspects need improvement. There is heavy usage of words like ”we” used. Also, some statements are made by (Wu et al., 2016) that are not cited to supported their claim. This is evident when its said ”As deep learning for video analysis is an emerging and vibrant field...” (According to whom?).

Subject independent facial expression recognition with robust face detection using a convolutional neural network by Matsugu et al. illustrates the difficulties that may arise when performing facial recognition. They highlight the problems that may occur in terms of being able to handle variability of subject faces, and certain angles. Their approach to this problem is addressed by implementing a rule based algorithm that analysis the results given from the CNN. Furthermore, their model is designed to be segment and be trained on specific facial features instead the face as a whole. Their model proves to be a success as they score an accuracy of 97.6 percent, also they do not require a second CNN working concurrently to achieve similar results as other models have done so previous, which can be cost effective. Some similarities arise between this paper and the proposed project, they both use sentiment analysis and require the ability to handle a wide range of variability. This proves beneficial

to the proposed project as it provides inspiration to use a rule based algorithm for determining emotions. Even though this paper is well written, there are some issues. In certain parts there are abbreviations to words given without the full word being given prior which can cause confusion to the reader. For example: "FP neurons". In addition to this, their model is specific for smiling faces and doesn't accommodate for other emotions, which should be at least provided in a further work section.

Arthmann and Li's paper titled *Neuromarketing: The Art and Science of Marketing and Neurosciences Enabled by IoT Technologies* is a promising insight to the field of neuromarketing. They recognize the association of online shopping and loss of sale for retail stores, and give example of how neuro-linguistic programming and IoT technologies can be used as a combination to understand their customers sentiment. Furthermore, it's heavily argued that the integration of AI and machine learning will evolve this concept to understand the thoughts of consumers and further tackle loss in productivity. This proves relevant to the proposed project as it is very similar. They both aim to gain a deeper understanding of human sentiment that may not be expressed verbally. Although this is possibly the most interesting paper of this review, it is severely lacking citations. Also, there are a lot of assumptions brought forward with no clear indication as to how this knowledge is known. Additionally, it is also assumed that people will adopt these "always on" IoT devices and agree for their physical aspects to be used for consumer targeted marketing.

Lopes et al. go into great detail of the past implementations of facial recognition systems. They review a number of methodologies to reinforce the idea of data augmentation with their own work. It is stated that due to the lack of datasets that are publicly available and overlapping of data, much of the accuracy results that are given by past works may be inaccurate (Lopes et al., 2017). Furthermore, they enable the reader to grasp a deep understanding to not only how convolutional neural networks operate, but how the common methods of implementation should be done. This is evident when explaining their system's implementation which was broken into two segments: The training phase and testing phase. Additionally, they give a clear and comprehensible description of the methods chosen and



why they were taken. For example: using synthetic sample generation, rotation, cropping, down sampling, and normalisation. This work proves relevant to the proposed project as it gives an explanation of how a segment of the project should be made. Also, Lopes et al. make it very clear that the CK+ dataset is the most suited dataset for facial and sentimental expression recognition.

## **2.7 Conclusion of Literature Review**

In conclusion, the five papers reviewed topics in relation to convolutional neural networks, facial sentiment analysis, emotion detection and current applications in the real world. Relevant aspects include detecting employees emotions in a work environment, implementing a score fusion algorithm for achieve a summation of two sentiment detecting technologies, the idea of segmenting facial features while training to accommodate for variability and the notion using human emotion understanding for a business solution. Lastly, an indication to what data that should be utilized is obtained throughout the range of papers that have been reviewed.

# Chapter 3

## Analysis

The following section will give an analytic insight to the steps required for designing the proposed software in relation to a business use case, Identify some of the problems that may arise and outline objectives for implementation.

An article by Hague and Hague for the B2B international states that understanding a customers satisfaction rate is important as it can show where is the business is doing well or where it needs improvements. It also can give indication to business owners where a further staff training is needed, or how there may be a need for cultural change. In light of this, seeing your business through the eyes of provides understanding of its downfalls. This in turn prevents more customer churn and can prove to be financially beneficial (Hague and Hague, 2017). They state that the downfalls of customer service surveys is the that they survey must "ask the right question of the right person". Hague and Hague give two reasons why this is difficult: they may know what to ask the customer in respect to their specific interaction with the business, also they may not have the contact details of the customer to further inquire into satisfaction of the individual. A main reason why customer surveys result may be misleading is the fact that an average of only 10% percent of online surveys sent to customers are responded to, meaning that these results are not representative of an entire customer base (Willott, 2011) . Furthermore, surveys are described by Hague and Hague as a "snapshot at one point in time", and do not accurately represent the feelings of the customer during transaction process, and measuring the satisfaction of customers should

”be a continuous process”.

Kirkpatrick describes the use of artificial intelligence technologies in customer service as a beneficial factor for businesses, as it can automate the tasks that are ”too time consuming”. In consideration of the aforementioned problems in the customer service sector, the proposed technologies aims to eliminate the process of requesting customers to fill out surveys, and in turn reduce cost of marketing teams investing funds onto campaigns that me be driven by misleading or inaccurate data.

In order to achieve the end goal of this project, a number of factors should be addressed. These consist of the type of model that should be implemented, what machine learning library should be used, which programming language is preferable for this project, how may the model be trained and what are the options for deployment. For these problems to be address, the project analysis shall be decomposed in to four components: The Machine Learning model, Data preparation, Training and Hosting/Deployment

## **3.1 The Machine Learning Model**

### **3.1.1 Algorithm**

When dealing with image a classification problem such as facial sentiment analysis, there are two machine learning algorithms that deem worthy. Deep Belief Networks (DBN’s) were developed as an alternative to back propagation by Geoff Hinton. The idea behind DBN’s involves stacking of a Restricted Boltzmann Machines in order to train in a ”greedy manner” (Hinton, 2017). This model learns to excerpt hierarchical representations of training data. The other algorithm is convolutional neural networks (CNN’s). CNN’s consist of a number of convolution and sub-sampling layers. It is then followed by a dense or otherwise known as a fully-connected layer. It also uses back-propagation to adjust the weights of the network during training (Gupta et al., 2017).

### 3.1.2 Programming Language

There are many programming languages that can be used for machine learning and artificial intelligence. Brownlee describes the quest for choosing a programming language to be rather difficult, as the choice should be tailored around "your own requirements" and entirely depends on the project being undertaken. However, a number of opinions on popular machine learning languages are given. Firstly, MATLAB/Octave is described as being "excellent" for dealing with matrices and linear algebra (Brownlee, 2016a). The R language is shown to be a main choice for machine learning as it provides many machine learning algorithms and it is a great choice for developing advanced models. Brownlee states, although it is a good choice, it is seen as having a hard learning curve at first use. Python is said to be competition to MATLAB and R as its large range of libraries and abilities is widely used in the field of data science (Brownlee, 2016a). An article by Verma sheds light on most popular programming languages in terms of job listings for the year of 2016. Coming in first place was Python, as its popularity is unmatched by any other programming language. Secondly is Java, and after that R is seen to be the third most popular (Verma, 2017). See Figure 3.1 for graphical representation of Verma's findings.

In light of the proclamations made above, it is clear that the preferred languages should be easy to learn, are backed by a large community, and is capable of complex mathematical operations. Also, it is desirable that the languages chosen for the proposed language is fast and efficient because of the complex computations that will occur during training.

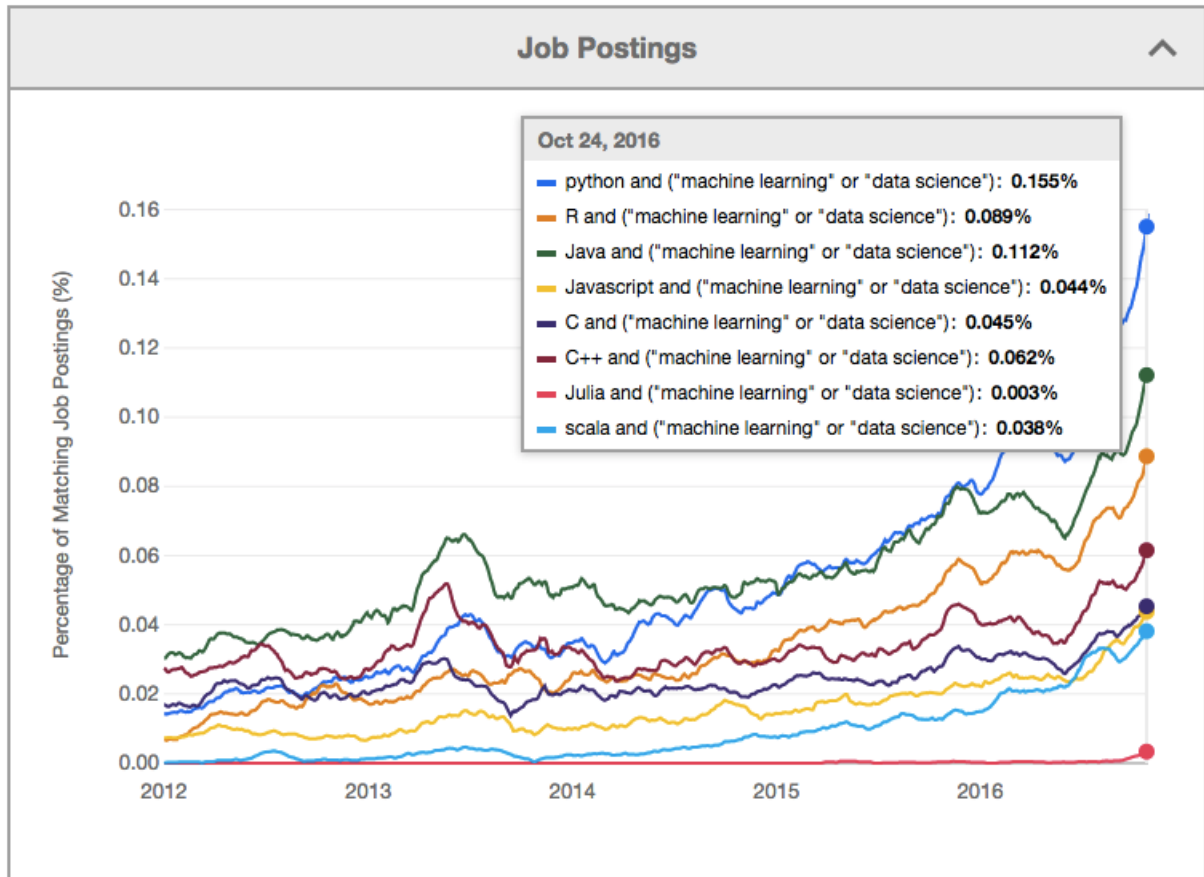


Figure 3.1: Graph of Machine Learning or Data Science Languages for 2016 by Verma (2017)

### 3.1.3 Machine Learning Library

To speed up the process of implementation, a machine learning (ML) should be used. This will speed up the process of the project as many ML libraries have API's to help eliminate tedious tasks. Also these can be use to reduce the complexity of the length of the program (Jain, 2017). TensorFlow is an open source ML library developed by the Google Brain team. It can be run on CPUs, GPUs and mobile platforms. For machine learning algorithms, such as gradient descent, TensorFlow provides automatic differentiation which can prove to be very advantageous in comparison to other ML libraries. It also supports multiple languages such

as Python, Java and C++ and Go (Jain, 2017). Graph visualisation is also supported. Multi-threading is also achievable with the use of TensorFlow (Jain, 2017). Higher level wrapper libraries such as Keras can use back-end API's such as TensorFlow provide easier implementation with minimalistic coding but also giving the same efficiency and accuracy (Brownlee, 2016b). Scikit Learn is an open source ML library built on top of other libraries such as Matplotlib, SciPy and NumPy. A good feature that Scikit Learn incorporates is the ability of evaluating, chaining and adjusting model hyper parameters (Jain, 2017). other ML libraries such as Caffe focuses on speed and modularity and is mainly utilized for convolutional neural networks and computer-vision. Another selling point for Caffe is its pre-trained models that do not require any coding or training and it supports GPU and CPU computations. A disadvantage of this library is that it's specifically designed for application implementation, not for research and development (Jain, 2017)

## **3.2 Data Preparation**

For this project, the machine learning model will require a large image dataset. This dataset will be used to train the model on certain facial feature expressions with supervised learning, therefore a labelled dataset is needed. This phase in implementation is the most important part of the project, as the model only as good as the data you train it with.

### **3.2.1 Public Datasets**

As seen in the literature review, there are many datasets publicly available for use on the Internet that provide labelled data. For example, the CK+ dataset was used by Lopes et al. used labelled data to train their model. The dataset included facial expressions such as sadness, surprise, happiness, anger, fear and disgust, which would be are emotional expressions that are ideal for the proposed project.

### 3.2.2 Creating a Dataset

Creating a dataset is another option for training the model. A number of images could be batch downloaded from image search engines such as Bing and Google, and split into training and testing sets. As for labelling, the name of the folder the type of picture is in will serve as the label for validation.

## 3.3 Training

Training is a vital part of the project and will be the most time consuming. The model will use the chosen data preparation method and use that data as input to the model. There are two ways the model can be trained: Locally and on an Infrastructure as a Service (IaaS).

### 3.3.1 Training Locally

Training locally involves running code on your own machine to train the model. A main advantage to this is that it is low cost. There are no fees in this approach besides the electricity used by the machine. However, this approach may be harmful to the machine as the high rate of computations produces a lot of heat, which may prove to be detrimental to your hardware. Also, should the setup not have the sufficient hardware, such as a high performing graphical processing unit (GPU), the training elapse time will be significantly larger.

### 3.3.2 Cloud Training

Training in the cloud involved pushing the model code and dataset to a cloud platform service. These infrastructures provide dedicated hardware for machine learning training such as high performance GPU's and CPU's. They usually provide container environments to run the training. For example, should a model be implemented in Keras or Caffe, the platforms have dedicated runtime environments with these libraries pre-installed. Notable platforms for machine learning training are TensorPort, FloydHub and Amazon's AWS. Although it is

faster to train a model on one of these platforms, they can be expensive, as the free tiers only provide a certain amount of server usage time.

## 3.4 Hosting/Deployment

Deployment of the trained model is required to serve it as a web service. This model should be hosted as an API to enable use from devices besides the machine it was trained on. The API will require a platform that containerizes and supports the machine learning libraries it was used to train with. Additionally, the API should be lightweight and fast for accurate classifications. There are many Platforms as a Service (PaaS) that can be used for this project, notably Heroku, AWS, Azure and Google App Engine all provide a free tier basis for deployment.

## 3.5 Concluding Objective

In conclusion, this section outlines the objectives for the proposed project sequentially from the initial steps to the end product. The objectives are as follows:

- Build a convolutional neural network for facial expression classification.
- Train the model on a large image data set.
- Save the trained model and deploy it to a Python API.
- Develop a web application that is enabled to record a user's facial expressions and send snapshots of the users face to the Python API in intervals.
- Use image pre-processing API on the web cam images for more accurate predictions.
- Use tone analyzer API for voice sentiment recognition.
- Weigh scores from the train model and tone analyzer accordingly and store them in a database



# Chapter 4

## System Design

The following chapter will explain the design process for the prototype for the project. The components of the project is broken into four parts: Machine Learning Algorithm on what algorithm and ML library was used, Data Preparation for insight into the data utilised, Training for explanation on the approach for training the model and Deployment for the process in which the model was deployed for production. Lastly, the overall system architecture will be explained.

////\*\*\*\*\* This will be later updated for the final design \*\*\*\*\*///

### 4.1 Machine Learning Model

For designing the model, the Keras machine learning library with the TensorFlow backend was used in order to easily create a convolutional neural network. The network consists of two convolution layers, two MaxPooling layers and a fully-connected layer. The model has two outputs nodes that provide classification of two emotions: happy or sad. The model used supervised learning, with labels for the images it was training on, which will be brought forward in the following section.

## 4.2 Data Preparation

As stated by Lopes, de Aguiar, Souza, and Oliveira-Santos that there is a scarcity of public datasets with for facial expression images, therefore a dataset was created for created by bulk downloading images from Google Images on searches such as "Happy Human face" and "Person Frowning". These images where split into a 30/70 split. The training data consisted of 70% of the images and the remaining 30% was used in testing. Within these these training and test sets are folders titles **happy** and **sad**. These folder names where used as labels during the training phase of the prototype.

## 4.3 Training

Due to insufficient hardware, the approach of cloud training was needed in order to train the model. The Python code and the dataset were pushed to a containerized TensorFlow environment on the FloydHub cloud platform. The output of the trained weights is saved to a *.h5* file and the trained model is JSON file.

## 4.4 Deployment

In order to deploy the trained model to production, a Python API was developed using the Flask package, which provides many libraries to run python code on a server. The API is hosted on Heroku PaaS as they provide an easy-to-use and free basice tier service for hosting Python apps. The trained weights and model files are also deployed with the API. These are read in and initialised from their respective files when the API successfully build and deploys. The API accepts 'POST' requests and takes the image as input. The image is then fed into the model and a classification of **Happy** or **Sad** is made, this is then returns to the original sender.

Secondly, a Node.JS web app was developed in order to display how this model can be used in production. The application was also deployed to heroku. The application uses

the clients webcam in order to record their face. This takes snapshots of the users face and sends the images to the Python API in two second intervals. The application then receives the response from the API and displays it to the screen in the format of '*Emotion Detected: Happy*' for example. The web app is simple in design but more features will be added in the final implementation, which will be explained in the system architecture section.

## 4.5 Concluding Remarks of System Architecture Design

Figure 4.1 illustrates a logical topology of final project. The primary component such as a web application that records the users face and the model deployment have already been achieved.

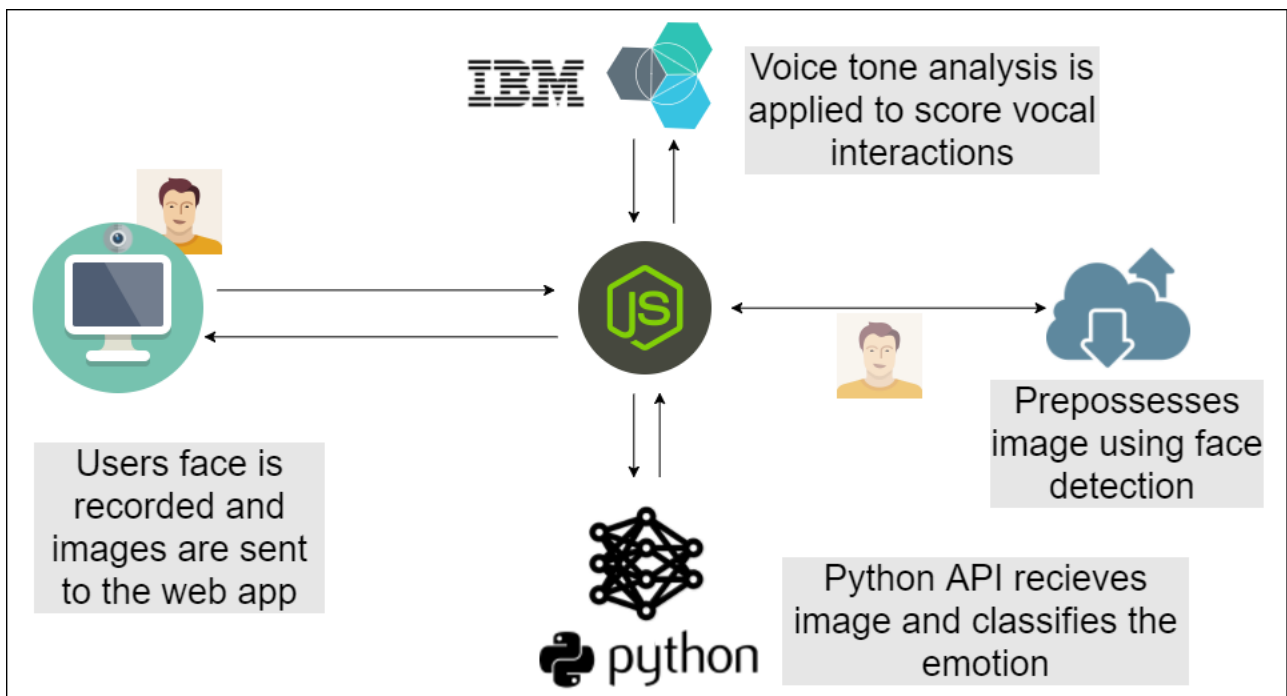


Figure 4.1: System architecture of the prototype

However, further functionalities such as image preprocessing of the captured images is needed for more accurate classifications. Also, a tone analysis feature should be implemented to reach the over all goal of the project. In conclusion, it can be noted that 1/2 of the project is achieved within this prototype.

# **Chapter 5**

## **Implementation**

### **5.1 Implementing the Machine Learning Model**

### **5.2 Deploying the Trained Model**

### **5.3 Implementing the Node.JS Application**

#### **5.3.1 Database Integration**

### **5.4 Result Weighting Algorithm**

### **5.5 Integrating Third Party Services**

### **5.6 Conclusion**

# **Chapter 6**

## **Testing and Evaluation**

### **6.1 Testing Model with Testing Data**

### **6.2 API Testing**

### **6.3 Testing Model with New Data**

# **Chapter 7**

## **Results and Discussion**

### **7.1 Results**

#### **7.1.1 Model Accuracy**

#### **7.1.2 Usability**

### **7.2 Discussion**

# **Chapter 8**

## **Conclusion and Further Work**

### **8.1 Conclusion**

#### **8.1.1 Further Work**



# Bibliography

Ahmad Abdulkader, Aparna Lakshmiratan, and Joy Zhang. Introducing deeptext: Facebook's text understanding engine, 2017. URL <https://code.facebook.com/posts/181565595577955/introducing-deeptext-facebook-s-text-understanding-engine/>.

Christopher Arthmann and I-Ping Li. Neuromarketing the art and science of marketing and neurosciences enabled by iot technologies, 2017. URL [https://www.iiconsortium.org/pdf/2017\\_JoI\\_Neuromarketing\\_IoT\\_Technologies.pdf](https://www.iiconsortium.org/pdf/2017_JoI_Neuromarketing_IoT_Technologies.pdf).

Nikhil Bhargava and Manik Gupta. Application of artificial neural networks in business applications. *GEOCITIES*, 1:3–4, 2017.

Jason Brownlee. Best programming language for machine learning, Sep 2016a. URL <https://machinelearningmastery.com/best-programming-language-for-machine-learning/>.

Jason Brownlee. Introduction to python deep learning with keras, Oct 2016b. URL <https://machinelearningmastery.com/introduction-python-deep-learning-library-keras/>.

Adit Deshpande. A beginner's guide to understanding convolutional neural networks. <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>, 2016. [20th July 2016].

Macedo Firmino, Antnio H Morais, Roberto M Mendoa, Marcel R Dantas, Helio R Hekis, and Ricardo Valentim. Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects, Apr 2014. URL <https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-13-41>.

Dishashree Gupta, Pranjal Srivastava, Shubham Jain, and Faizan Shaikh. Architecture of convolutional neural networks (cnns) demystified, Jun 2017. URL <https://www.analyticsvidhya.com/blog/2017/06/architecture-of-convolutional-neural-networks-simplified-demystified/>

Paul Hague and Nick Hague. Customer satisfaction survey: How to measure satisfaction, 2017. URL <https://www.b2binternational.com/publications/customer-satisfaction-survey/>.

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron C. Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *CoRR*, abs/1505.03540, 2015. URL <http://arxiv.org/abs/1505.03540>.

Salakhutdinov Hinton. Deep belief networks, Oct 2017. URL <http://deeplearning.net/tutorial/DBN.html>.

Rashmi Jain. 7 powerful programming languages for doing machine learning, Jan 2017. URL <http://blog.hackerearth.com/powerful-programming-languages-for-machine-learning>.

Keith Kirkpatrick. Ai in contact centers: Artificial intelligence technologies are being deployed to improve the customer service experience. *Communications of the ACM*, 60(8):18 – 19, 2017. ISSN 00010782. URL <http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=bth&AN=124418580&site=eds-live>.

Andr Teixeira Lopes, Edilson de Aguiar, Alberto F. De Souza, and Thiago Oliveira-Santos. Facial expression recognition with convolutional neural networks: Coping with few data

- and the training sample order. *Pattern Recognition*, 61(Supplement C):610 – 628, 2017. ISSN 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2016.07.026>. URL <http://www.sciencedirect.com/science/article/pii/S0031320316301753>.
- Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, and Yuji Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5):555 – 559, 2003. ISSN 0893-6080. doi: [https://doi.org/10.1016/S0893-6080\(03\)00115-1](https://doi.org/10.1016/S0893-6080(03)00115-1). URL <http://www.sciencedirect.com/science/article/pii/S0893608003001151>. Advances in Neural Networks Research: IJCNN '03.
- R. Subhashini and P.R. Niveditha. Analyzing and detecting employee's emotion for amelioration of organizations. *Procedia Computer Science*, 48(Supplement C):530 – 536, 2015. ISSN 1877-0509. doi: <https://doi.org/10.1016/j.procs.2015.04.131>. URL <http://www.sciencedirect.com/science/article/pii/S1877050915006407>. International Conference on Computer, Communication and Convergence (ICCC 2015).
- Adarsh Verma. Most popular programming languages for machine learning and data science, Jan 2017. URL <https://fossbytes.com/popular-top-programming-languages-machine-learning-data-science/>.
- Lindsay Willott. Customer satisfaction surveys - 5 reasons why they don't work, 2011. URL <https://www.customerthermometer.com/customer-surveys/5-reasons-online-customer-satisfaction-surveys-dont-work/>.
- Zuxuan Wu, Ting Yao, Yanwei Fu, and Yu-Gang Jiang. Deep learning for video classification and captioning. *CoRR*, abs/1609.06782, 2016. URL <http://arxiv.org/abs/1609.06782>.

# **Appendices**

## **Appendix A Code Snippets**

## **Appendix B Screen shots**

## **Appendix C Neural Network Diagram**