**CAPTIONED**

A CAPSTONE PROJECT REPORT

BY

PANAV TYAGI (E17CSE184)

A close up of a sign

Description automatically generated

SUBMITTED TO

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING BENNETT UNIVERSITY

GREATER NOIDA, 201310, UTTAR PRADESH, INDIA

in

*Partial fulfillment of the requirements*

*for the degree of*

BACHELOR OF TECHNOLOGY

DECEMBER 2021

# CERTIFICATE

This is to certify that the capstone project report entitled “**Captioned**” is being submitted by **Mr. Panav Tyagi** (Enroll. No. E17CSE184)) to the Department of Computer Science Engineering, Bennett University, Greater Noida, in partial fulfillment of the requirements for the award of Degree of Bachelor of Technology. It is an original research work carried out by him/them as the 8th Semester course from August 2021 to December 2021.

The report has fulfilled all the requirements as per the regulations of this institute and has reached the standard needed for submission. The results embodied in this capstone project report has not been submitted to any other university or institute for the award of any other degree or diploma, degree elsewhere.

**Panav Tyagi**

(Enroll. No. E17CSE184)

This is to certify that the above statement made by the candidate(s) is correct to the best of my knowledge.

**Dr. Suneet k. Gupta** (Capstone Project Instructor)

Assistant Professor

Computer Science Engineering Department

Bennett University, Greater Noida, INDIA

# DECLARATION

We hereby declare that the work which is being presented in the report entitled “Captioned”, in partial fulfillment of the requirements for the Bachelor of Technology in Computer Science and Engineering is an authentic record of my/our own work carried out during the period from August 2021 to December 2021 at Department of Computer Science and Engineering, Bennett University Greater Noida.

The matters and the results presented in this report has not been submitted by me/us for the award of any other degree elsewhere.

Signature of Candidate

Panav Tyagi

(Enroll. No. E17CSE184)

# ACKNOWLEDGEMENT

I would like to take this opportunity to express my/our deepest gratitude to my/our mentor, **Dr. Suneet K. Gupta** (Assistant Professor, CSE Bennett) for guiding, supporting, and helping us in every possible way. I was extremely fortunate to have him as my mentor as he provided insightful solutions to problems faced by me/us thus contributing immensely towards the completion of this capstone project. We would also like to express my sincere gratitude towards **Dr. Deepak Garg,** (HOD, CSE Bennett) for providing us constant support and all the necessary resources needed to complete this project.

Signature of Candidate

Panav Tyagi

(Enroll. No. E17CSE184)

TABLE OF CONTENTS

[TABLE OF CONTENTS iv](#_Toc56423391)

[LIST OF TABLES vi](#_Toc56423392)

[LIST OF FIGURES vii](#_Toc56423393)

[ABSTRACT viii](#_Toc56423394)

[1. INTRODUCTION 1](#_Toc56423395)

[1.1. Problem Statement 1](#_Toc56423396)

[2. Background Research 2](#_Toc56423397)

[2.2. Proposed System 2](#_Toc56423398)

[2.3. Goals and Objectives 3](#_Toc56423399)

[3. Project Planning 3](#_Toc56423400)

[3.1. Project Lifecycle 3](#_Toc56423401)

[3.2. Stakeholders 5](#_Toc56423402)

[3.3. Project Resources 5](#_Toc56423403)

[3.4. Assumptions 6](#_Toc56423404)

[4. SYSTEM ANALYSIS AND DESIGN 6](#_Toc56423405)

[4.1. Overall Description 6](#_Toc56423406)

[4.2. Users and Roles 7](#_Toc56423407)

[4.3. Design diagrams & Flow Charts 8](#_Toc56423408)

[4.3.1. Algorithm flow Diagrams 8](#_Toc56423409)

[4.3.2. User Flow Diagram 8](#_Toc56423410)

[5. User Interface 9](#_Toc56423411)

[5.1. UI Description 9](#_Toc56423412)

[5.2. UI Mockup 10](#_Toc56423413)

[6. Project Closure 11](#_Toc56423414)

[6.1. Goals / Vision 11](#_Toc56423415)

[6.2. Delivered Solution 11](#_Toc56423416)

[6.3. Remaining Work 12](#_Toc56423417)

6.4. [References 12](#_Toc56423418)

LIST OF TABLES

Table Page

[Table 1: Goal and Objectives 3](#_Toc56339789)

[Table 3: Stakeholders 5](#_Toc56339790)

[Table 4: Resources 6](#_Toc56339791)

[Table 5: Assumptions 6](#_Toc56339792)

[Table 12: User Roles 7](#_Toc56339793)

LIST OF FIGURES

Figure Page

[Figure 1: Complete Algorithm. Text Generation and Emotion Analysis Algorithm 8](#_Toc56339743)

[Figure 2: User Flow Diagram 9](#_Toc56339744)

[Figure 5: Website Upload Screen 10](#_Toc56339745)

[Figure 6: Results Screen 11](#_Toc56339746)

ABSTRACT

Captions are a significant part of any social media post. They are basically a tool that provide you an insight into the image in less amount of words. In this project I have implemented an algorithm which can be used to generate text or captions based on the emotion that is detected in the faces present in the uploaded image. The project has two parts – An emotion detection algorithm and a text generation algorithm. The emotion detection algorithm has been created using CNN. The dataset that was used to create this model was obtained by scraping images from Instagram. I also used the FER-2013 dataset. The resultant model gave us an accuracy of 65%. I validated my model via the scraped images from social media.

The second part of this project, is the text generation model that was created with the help of a python library for gpt-2, named gpt-2-simple. I scraped text for different emotions from the web and various social media sites. The text was cleaned using standard corpus cleaning methods and finally it was consolidated to be used as input for the text generation model. The gpt-2-simple module was then trained on my corpus and the generated text was both grammatically and contextually correct. Furthermore, the project is available on two interfaces – a website and an android application. The website was created using Flask and the application was developed using Flutter. The project has been deployed on Heroku. All my code, demos and datasets will be open sourced and made available on GitHub.

1. INTRODUCTION

As they say, “a photograph is worth a thousand words”. In any case, this is just obvious if your photograph's organization draws the consideration of the watcher to its subject. In this sense, a caption can be considered as a composition instrument. One approach to attract your watcher's eye to your photograph's subject is to utilize a photography creation rule, for example, the Rule of Thirds. Likewise, a caption can attract a watcher's eye to a significant part of your photo. What separates a caption from other creation apparatuses is it can even cause to notice data that is absent in the photograph!

Captions are a powerful asset that can be helpful in gathering likes, comments, and can even get an online viewer to connect with you emotionally. Moreover, if your post gets enough traction/engagement, it can land up on the *explore page* which helps gain a bigger social media following. Most importantly, captions create an enormous opportunity to establish a more grounded relationship with your audience. An extraordinary social media post caption will leave a viewer/surfer speechless and the more somebody spends reading your caption and connecting with your post, the better your post will rank in accordance with the Instagram algorithm.

Therefore, as part of a generation obsessed with posting on social media, I decided to create a software that can help people generate captions based on the mood displayed in their images. I likewise find automated caption generation intriguing on the grounds that it caters to the needs of a content creator or a consumer who likes posting on social media.

* 1. Problem Statement

Captions play a significant role in increasing your engagement on social media platforms. An interesting caption can help increase your outreach, and henceforth increase your social media following. A solid caption paired with your post can spark the reader's attention in a full text story. A caption, in its essence, is its own mini story. The more attention your post receives on a social media platform, the better is it is ranking on the platform, that would in turn help you increase your online following.

Avid social media users must go through the hassle of finding a perfect caption for their image every time they post anything. They go through blogs, search multiple things on google just to find a perfect caption for their image. I thought that what if there is a tool which can suggest a user some captions based on the emotion in their image. This would, to an extent, reduce the hassle of brainstorming and finding a good caption. Therefore, I decided to create a caption generator as our capstone project.

1. Background Research

Before choosing this subject as my semester venture, I went through different research papers and articles to check if my theme is attainable and gives the ideal outcomes or not. Various individuals had various methodologies and results be it regarding exactness or accuracy of classification of emotions or the attainability of the methodology. Following are the research papers that I went through before starting to work on my project:

[This](https://arxiv.org/pdf/1502.03044.pdf) paper deals with generating captions for images using neural attention. The paper suggests a state-of-the-art approach using BLEU and METEOR metric on 3 benchmark datasets. Following is an excerpt from the paper –

*“Inspired by recent work in machine translation and object detection, we introduce an attention-based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation techniques and stochastically by maximizing a variational lower bound. We also show through visualization how the model can automatically learn to fix its gaze on salient objects while generating the corresponding words in the output sequence. We validate the use of attention with state-of-the-art performance on three benchmark datasets: Flickr8k, Flickr30k and MS COCO.”*

Another paper that I went through can be found [here](https://arxiv.org/abs/1509.04942). The authors, in this paper, suggest a method of generating captions for a given image using the LSTM model. A passage from the paper gives the idea of the paper:

*“In this work we focus on the problem of image caption generation. We propose an extension of the long, short term memory (LSTM) model, which we coin gLSTM for short. In particular, we add semantic information extracted from the image as extra input to each unit of the LSTM block, with the aim of guiding the model towards solutions that are more tightly coupled to the image content. Additionally, we explore different length normalization strategies for beam search to avoid bias towards short sentences. On various benchmark datasets such as Flickr8K, Flickr30K and MS COCO, we obtain results that are on par with or better than the current state-of-the-art.”*

* 1. Proposed System

My task is to make and incorporate a profound learning NLP model to create subtitles for web-based media photographs utilizing notion investigation of the picture combined with face acknowledgment. The model will take in the picture that the client needs to transfer to his/her feed and would recommend A.I. produced subtitles for it dependent on the mind-set of the photo. I would utilize web-scrubbers to really assemble information for preparing the content age model which would scratch different Instagram and twitter pages gathering different sorts of inscriptions and jokes. The informational index will require substantial preprocessing to prepare the content age model successfully. The dataset will experience a period of opinion examination grouping the sentences and subtitles into 4-5 general classes that normally are related with a picture on an online media stage.

* 1. Goals and Objectives

Table : Goal and Objectives

|  |  |
| --- | --- |
| **#** | **Goal or Objective** |
| 1 | Make the system extensible – future updates like xxx can be done easily |
| 2 | Make the system easy to support – provide good documentation, configuration/build files, administrator’s manual |
| 3 | Make the system very easy to use – users would agree that minimal to no training is needed |
| 4 | Build a prototype that demonstrates the user interface by xx/xx/xx - in order to get early feedback from the customer/users |
| 5 | Have fun working on the project |

1. Project Planning
   1. Project Lifecycle

The team followed an agile approach of development. The team created a timeline of 12 weeks and targeted to complete the project by the end of October. I gathered requirements and created a high-level development plan at the onset of the project and then implemented the gathered requirements over a span of 12 weeks. Following is my week wise plan that I followed during the project:

**Week 1 -** As opposed to mainstream thinking that the initial stretch of the project is troublesome in light of deciding a project and different things, the first week was genuinely simple as a result of a Eureka moment that we had at the point when the thought flew into our head. *For the remainder of the week went into exploring the plausibility of the thought.* We went through different blogs and research papers to see if anyone has worked on something similar before and find out different methods and ways to do the project.

**Week 2 -** *We started working towards the data collection part* simultaneously. Not only we started looking for data sources on the web but also decided to get data from social media sites by creating scrapers to scrape the public data available on such social platforms.

**Week 3 –** During this week, *we completed building our scraper and started collecting data from the online resources.* We built an Instagram scraper to collect images through hashtags. We also created a twitter scraper to collect quotes for emotions through various twitter accounts that we know post particular kinds of stuff. We collected around 3000 tweets that were further cleaned to be used as input for the model. We also used the facial emotion recognition dataset present on Kaggle for training our model for emotion detection from the image.

**Week 4 -** We started working towards building a model for emotion recognition from images. We *tried some pre-trained models available on GitHub* to see what kinds of results they give and studied them further to understand what algorithms have been used to create those models. We *started working towards creating a CNN model* which we thought will be best suited for a task like ours. We created the model and tried refining the model by tweaking some parameters and adding/reducing layers until we got the best results.

**Week 5 -** This week, we proceeded towards creating a next-generation model*. We started by creating some basic RNN and LSTM models.* But these models were not accurate enough as we were looking for results that were both grammatically and contextually correct. After some exploration on the web, we found a python library that was an implementation of the gpt-2 model by Open-AI. That library could be used to train the gpt-2 model on our own text and get results. After testing for a few days, we finally got what we wanted.

**Week 6 -** The main functionalities of our project, that is, the models were ready. Now it was *time to integrate both the models in a single project to achieve the desired outcome.* We decided to create a REST API for the models. Initially, we were of the opinion that we should create a single API endpoint for both functionalities and started working towards it but eventually, we realized that it would be better to create two separate endpoints for both the models as it would be beneficial for us while creating the app.

**Week 7 -** *While we were still working on the API, we started designing the front end* simultaneously. We created a minimal frontend for the website just for the sake of testing while we were still working on the main design. The API creation was completed. We tested the APIs using cURL and Postman and they were working fine. We also started documenting the progress of our project.

**Week 8 -** *We started working towards the design of the mobile application*. We also started working towards integrating a new feature in our application which included suggesting hashtags and emojis based on the generated captions.

**Week 9 -**During this week, *we worked towards the development of web and mobile applications*. While it as easier for us to create a web application since we have prior experience of working with web-based frameworks, creating the mobile application took some time due to lack of previous experience.

**Week 10 -** During this week, *we completed the web development part and a major chunk of the mobile application*. We then began with the testing of our application. After the alpha testing was done, we went into the beta testing phase where we gave the app to some of our friends and took their feedback regarding the user experience and functionality.

**Week 11 and 12 -** During these two weeks, *we completely focused on the deployment of the project on the cloud*. As soon as the applications were deployed, we got back to the drawing board to do testing to ensure that everything was working fine. We then sat down to write the final documentation of the project and started preparing for the submission.

* 1. Stakeholders

The stakeholders of the project are listed below.

Table : Stakeholders

|  |  |
| --- | --- |
| **Stakeholder** | **Role** |
| Dr. Deepak Garg | Sponsor (Head of Department) |
| Dr. Suneet K. Gupta | Mentor |
| Dr. Indrajeet Gupta | Instructor |
| Panav Tyagi | Team member |

* 1. Project Resources

The resources that were required to complete this project are listed as follows:

Table : Resources

|  |  |  |
| --- | --- | --- |
| **Resource** | **Resource Description** | **Quantity** |
| Storage Service | An online storage service to temporarily store the uploaded data. | 1 |
| Capstone Team | Team of students to create the tool as stated previously. | 2 |
| Dr. Suneet K. Gupta | My Mentor for the project to guide us and support us if we land into any issues while creating the tool. | 1 |
| Windows Workstation | As the team was familiar with windows, Windows workstations were required to create the project. | 2 |
| Android Phone | An android supported device was required to test the mobile application. | 1 |
| Internet connection | A stable internet connection was required during the development process to test the tool and also search for any queries on the web. | 2 |

* 1. Assumptions

Table : Assumptions

|  |  |
| --- | --- |
| **#** | **Assumption** |
| A1 | The capstone team and mentors will be able to meet face to face once a week. |
| A2 | The team members are constantly available for the duration of the project. |
| A3 | The mentors will be able to solve all our queries and doubts. |
| A4 | The end user is familiar with using an android application and web application. |
| A5 | The development test data provided will be sufficient to create an accurate prediction of user actions. |
| A6 | The pre-built libraries are accurate and provide correct results. |

1. SYSTEM ANALYSIS AND DESIGN
   1. Overall Description

My venture is to make and incorporate a profound learning NLP model to produce subtitles for web-based media photographs utilizing feeling investigation of the picture combined with face acknowledgment. The model will take in the picture that the client needs to transfer to his/her feed and would propose A.I. produced subtitles for it dependent on the mind-set of the photo. I would utilize web-scrubbers to really assemble information for preparing the content age model which would scratch different Instagram and twitter pages gathering different sorts of inscriptions and jokes. The informational collection will require substantial preprocessing in order to prepare the content age model viably. The dataset will experience a period of assessment examination arranging the sentences and inscriptions into 4-5 general classes that typically are related with a picture on an online media stage.

The mechanized caption generator spares the problem of an eager web-based media client who needs to experience a great deal of statements, articles, and other web-based media accounts looking for the ideal inscription for his/her post. Even after choosing the caption, there might be grammatical errors and language structure botches in the subtitle which debilitates the association between the post and the watcher. It will likewise guarantee that the subtitle consistently fulfills the guidelines making the entire “caption game” powerful and reliable all through concerning "quality". The subtitle generator really proposes inscriptions dependent on an investigation of the picture that implies that your subtitle and your post are consistently co-identified with one another and don't simply exist as two separate bodies.

* 1. Users and Roles

Table : User Roles

|  |  |
| --- | --- |
| **User** | **Description** |
| Social Media Influencers | Anyone who uses social media as an influencer can use this software to generate good captions for the content, they are posting to create a great impact on their followers. |
| Anyone who likes to post online | Anyone who likes to post on social media, likes to post a good caption with their image. Instead of brainstorming or searching on the web for good captions, they can use the software to generate captions based on the image. |

* 1. Design diagrams & Flow Charts
     1. Algorithm flow Diagrams

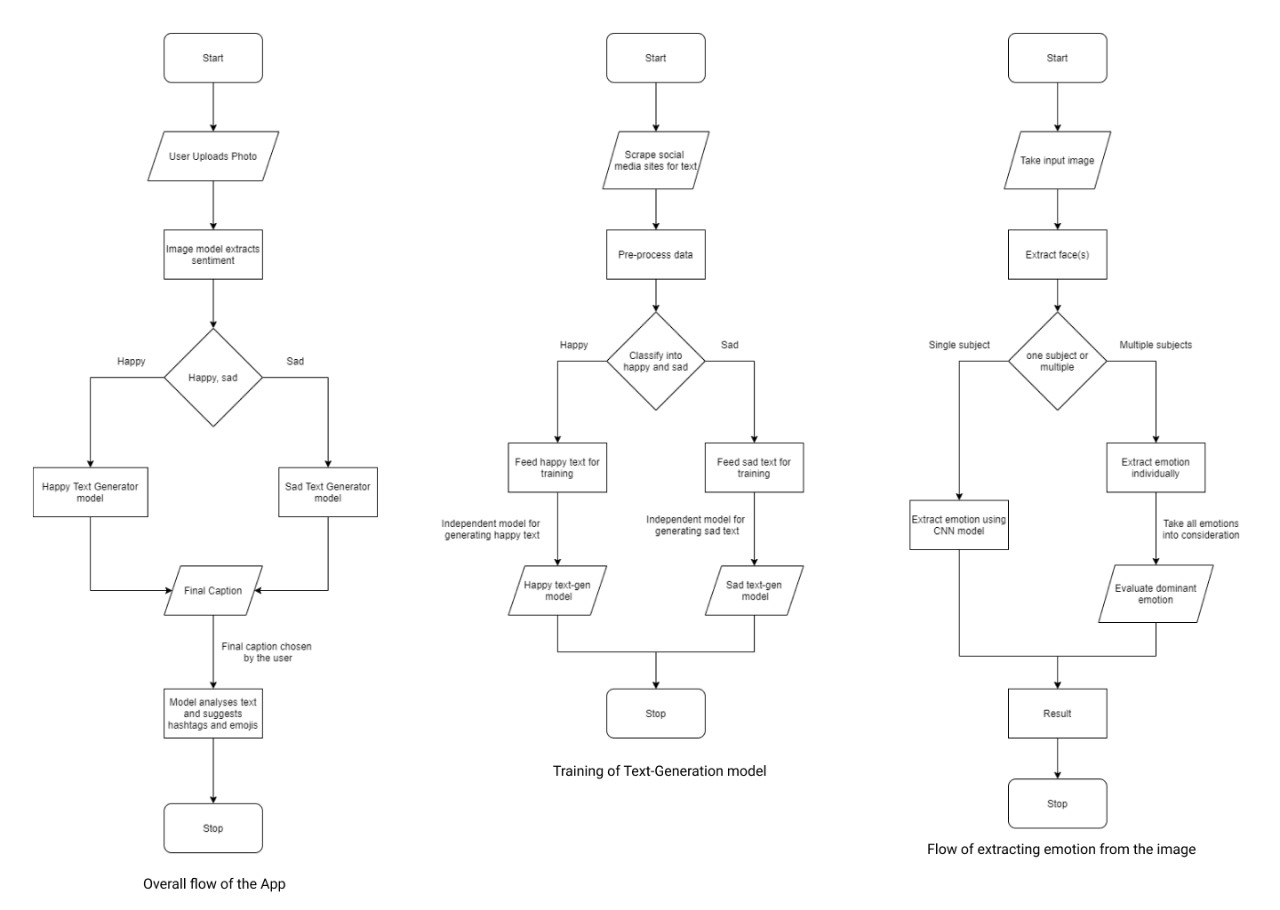


Figure : Complete Algorithm. Text Generation and Emotion Analysis Algorithm

* + 1. User Flow Diagram

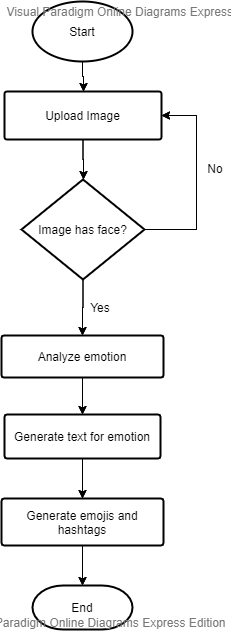


Figure : User Flow Diagram

1. User Interface
   1. UI Description

The web app and mobile app part of my project are still in their initial stages of development. The application interface is quite simple in nature. My application has straight forward U.I. implementations and concepts. All the user needs to do is upload the photo of his/her choice that is intended to be posted on the social media platform using the upload button. After that, the photo will be analyzed by my ML models and an AI-generated caption will be displayed which can be copied by clicking a button. Further, we would be adding in certain features that let you see the word count and some other insights regarding the caption and other related things.

* 1. UI Mockup

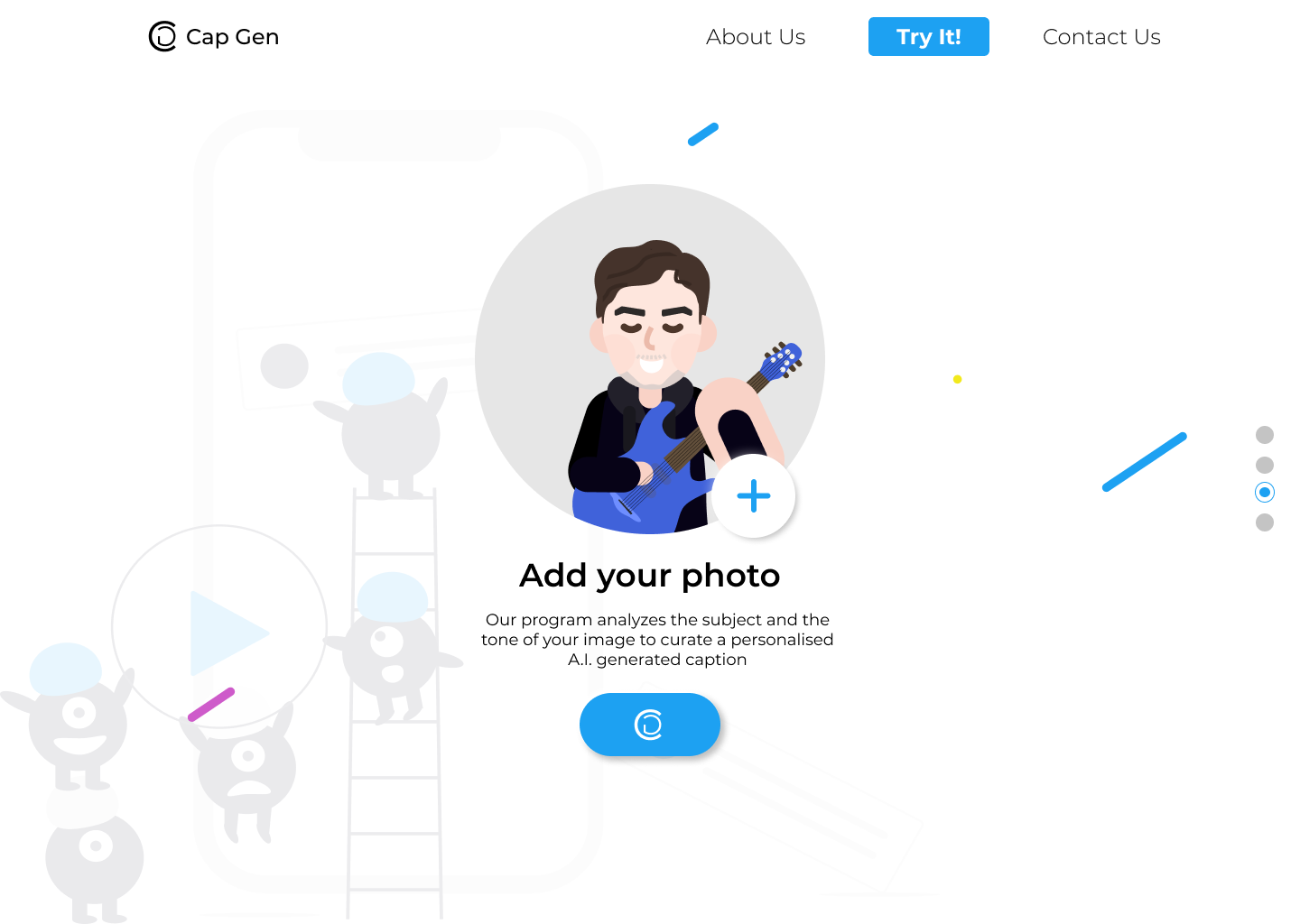
****

Figure : Website Upload Screen

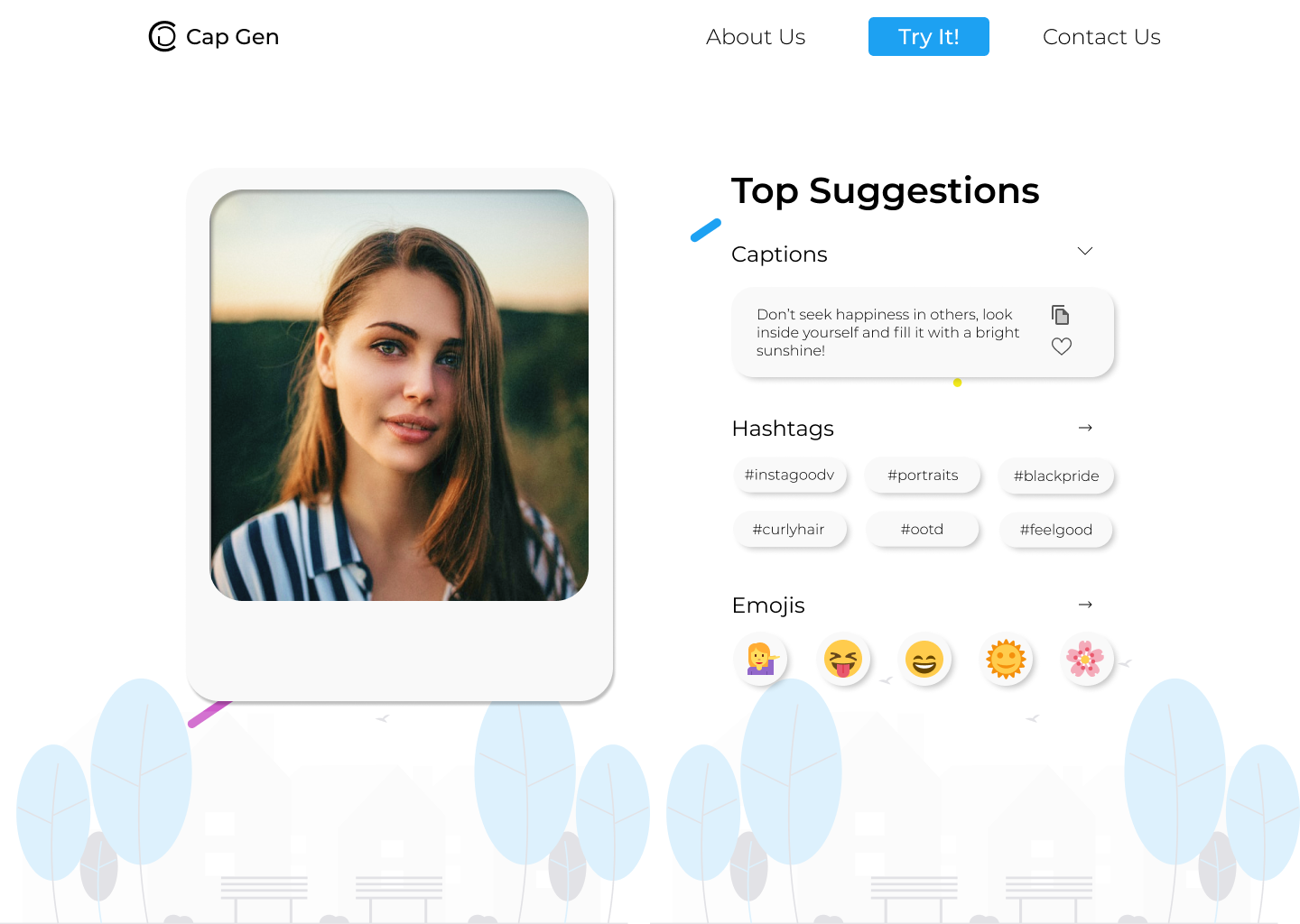


Figure : Results Screen

1. Project Closure
   1. Goals / Vision

My goal is to provide the users with AI-generated social media kits that are personalized just for them. We take in the image of the user and then using facial recognition techniques we can extract the emotions that are being portrayed in the image, in accordance with the actual mood of the portrait. My goal is to minimize the extra effort any avid social media user has to put in finding a perfect caption for their image. This project is basically a utility for people who like to post on social media and want a caption that perfectly captures the essence of their image.

* 1. Delivered Solution

The solution we delivered primarily consisted of a fully documented, fully featured web and android application. Both the applications have the functionality of uploading images after which captions are generated based on the emotion that is detected in the image. If the image does not have any human faces, an error is thrown which notifies you to enter an image with a human face. The delivered solution also generates the most suitable hashtags and emoticons based on the emotion detected in the image. Both the applications have an infinite loading feature which makes the experience in application seamless and increases the loading speed of the result. In case, no emotion is detected in the image, a neutral emotion is assigned, and random captions are generated.

* 1. Remaining Work

The project that we have created just now is just a prototype of the idea that we have in my mind. Therefore, the project just now generates captions only for “happy” and “neutral” emotion. We can incorporate more emotions like sad, angry, surprise, etc. in the coming time. Also, we can incorporate a feature where the user will be directly able to share the caption with the image on social media using the Instagram API. Also, we have thought of extending a project and creating a feature where even if there is no human face in the image, the model will be able to detect the mood of the image using properties like color, contrast, brightness, etc.

References

1. “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention” Kelvin xu, Jimmy Lei BA, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard S. Zemel, Yoshua Bengio
2. “Guiding the Long-Short Term Memory Model for Image Caption Generation” Xu Jia, Efstratios Gavves, Basura Fernando, Tinne Tuytelaars; Proceedings of The IEEE International Conference on Computer Vision (ICCV), 2015, Pp. 2407-2415
3. Bahdanau, Dzmitry, Cho, Kyunghyun, And Bengio, Yoshua. “Neural Machine Translation by Jointly Learning to Align and Translate”. ARXIV:1409.0473, September 2014.
4. “Challenges in Representation Learning: A Report on Three Machine Learning Contests” I Goodfellow, D Erhan, Pl Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, Dh Lee, Y Zhou, C Ramaiah, F Feng, R Li, X Wang, D Athanasakis, J Shawe-Taylor, M Milakov, J Park, R Ionescu, M Popescu, C Grozea, J Bergstra, J Xie, L Romaszko, B Xu, Z Chuang, And Y. Bengio. ARXIV 2013.
5. “Real Time Convolutional Neural Networks for Emotion and Gender Classification” Octavio Arriaga, Paul G. Plöger, And Matias Valdenegro. [arXiv:1710.07557](https://arxiv.org/abs/1710.07557) [cs.CV]