**VI Semester**

**Project Work - UE17MC652**

**Synopsis**

**Title**: Caption Generator

**Tools & Technology:**

* **Hardware Requirements**

|  |  |
| --- | --- |
| **Hardware** | **Specification** |
| **Processor** | **Intel(R) Core(TM) i7-6500U** |
| **Hard Disk** | **1 TB** |
| **Ram** | **8 GB** |
| **Keyboard & Mouse** | **Standard PS/2 Keyboard & ELAN I2C Filter Driver** |

* **Software Requirements**

|  |  |
| --- | --- |
| **Purpose** | **Tools & Technology** |
| **Frontend** | **HTML, CSS, Bootstrap** |
| **Backend** | **Flask 1.x** |
| **Language** | **Python 3.7** |
| **IDE** | **Jupyter Notebook, VS Code** |

**Abstract:** Caption Generatoris a Machine Learning Application which identifies the action portrayed in the given image. The generated caption will describe about the image that will say what kind of actions is taking place in it. This project involves computer vision and natural language processing concepts to recognize the context of an image and describe them in a natural language like English. The objective of the project is to build a working model of Caption Generator by implementing CNN with LSTM.

**Submitted by:**

|  |  |  |
| --- | --- | --- |
| **SRN** | **Name** | **Student signature with date** |
| PES1201702013 | Vijaykumar R Pai |  |

|  |  |
| --- | --- |
| **Internal Guide Name and Designation** | **Guide Signature with date** |
| Dr. S Thenmozhi |  |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**01** | | |
| From: 20**/01/2020** To: 25**/01/2020** | | |
| **Details of Work done**     1. Problem formulation 2. Decided the title for the project 3. Literature survey on the outline of methodologies to be used 4. Decided the tools and technologies required for the project 5. Decided the modules of the project. 6. Thought about the application in the real world 7. Prepared PPT for the title presentation | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**02** | | |
| From: 27**/01/2020** To: 01**/02/2020** | | |
| **Details of Work done**     1. Researched about CNN and LSTM. 2. Started learning about CNN and LSTM from tutorials. 3. Learned about the dependencies needed in the application. 4. Learned how to load images. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**03** | | |
| From: 03**/02/2020** To: 08**/02/2020** | | |
| **Details of Work done**     1. Imported the image dataset and its respective corpus. 2. Configured the GPU memory for training purposes. 3. Imported the required libraries. 4. Plotted few images and their captions from the dataset. 5. Cleaned captions for further analysis. 6. Cleaned captions for further processing. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**04** | | |
| From: 10**/02/2020** To: 15**/02/2020** | | |
| **Details of Work done**     1. Learning new algorithms and technologies   Detailed description: Learning about how CNN and LSTM can be implemented in my project. Learning the flask framework to design the web page.   1. Installing tools - Flask   Benefits and usage: Used for building UI for the ML Application.   1. Creating sample pages to test the functionalities provided by the flask framework. 2. Started creating a simple web page with HTML and CSS to show how the application will be once the ML Algorithm is integrated with it. 3. Plotting the top 50 words that appear in the cleaned dataset. 4. Adding start and end sequence tokens for each captions. 5. Loading VGG16 model and weights to extract features from the images. 6. Deleting the last layer of the model. 7. Started extracting features from the images. 8. Prepared PPT for first presentation. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**05** | | |
| From: 17**/02/2020** To: 22**/02/2020** | | |
| **Details of Work done**     1. Modified PPT – Added Research papers in literature survey. 2. Did comprehensive survey and research of Deep Learning for Image Captioning. 3. Added dynamic ability to display captions on uploading the image. 4. Started designing diagrams of the caption generator model, CNN and LSTM. 5. Added detailed Literature Survey with 3 existing systems. 6. Added Detailed Methodology. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**06** | | |
| From: 24**/02/2020** To: 29**/02/2020** | | |
| **Details of Work done**     1. Prepared myself for presentation. 2. Loaded VGG16 model and weights to extract features from the images. 3. Deleted last layer of the model 4. Started preparing the dataset for feature extraction 5. Extracted features from the images. 6. Did Principle Component Analysis to reduce the dimensions of the features and extract only important features from the images. 7. Plotted similar images from the dataset to see how many similar images exist in the bag of images. 8. Figuring out how to merge the images and captions for training. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**07** | | |
| From: 02**/03/2020** To: 07**/03/2020** | | |
| **Details of Work done**     1. Improvised feature extraction algorithm. Now able to store the extracted feature in pickle format. 2. Generated captions for each images and linked them to their images respectively. 3. Merged the images and the captions for training. 4. Found the maximum length of the caption. 5. Tokenizing the captions for further processing – Generated token of captions in pickle format. 6. Processed the captions and images as per the required shape by the model. 7. Splitted the training and test data – there is an array dimension error. Figuring out to rectify it. 8. Trying to figure out how to pass the extracted features to the LSTM model for generating captions for the images based on the actions depicted in it. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**08** | | |
| From: 09**/03/2020** To: 14**/03/2020** | | |
| **Details of Work done**     1. Loaded a single image to predict the probability across all output classes and retrieved the highest probability. 2. Generated VGG16 model summary. 3. Loaded the features stored in pickle format. 4. Loaded descriptions. 5. Converted the descriptions into whole description sequence model. 6. Generated word-by-word model to take out only important words for caption generation. 7. Merged the extracted features, token of words with the help of LSTM algorithm by using epochs = 20. 8. Need to take the best epoch model out of generated 20 that will be having least loss for better prediction. 9. Next need to make the caption generation a little more precise as now the prediction made by the model is not so accurate. 10. Need to evaluate model based on BLEU score. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**09** | | |
| From: 16**/03/2020** To: 21**/03/2020** | | |
| **Details of Work done**     1. Evaluated every model with BLEU Score to get model accuracy. 2. Ran epochs once again to generate model once again to improve the caption prediction and accuracy. 3. Manually running all the images in the dataset with all the 20 models to figure out the best model among them with the best accuracy for better prediction. 4. Able to run any images from the dataset and results are displayed immediately. 5. Trying to process images not listed in the dataset and get captions for those images. 6. Figuring out how to merge the generated ML model with web application. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |

**WEEKLY REPORT**

|  |  |  |
| --- | --- | --- |
| Week No:**10** | | |
| From: 23**/03/2020** To: 28**/03/2020** | | |
| **Details of Work done**     1. Refactored the whole project into a single class so that it will be easier to import it in flask during web development. 2. Imported packages: Flask, render\_template, request, redirect, url\_for, send\_from\_directory, flash, secure\_filename. 3. Created a folder called static in order to store the dataset in it for displaying the result during runtime. 4. Created a second webpage called caption.html to display result. 5. Starting creating web application to accept images such as png, jpg, jpeg. 6. Designing index page in order to accept any images whether it is in dataset or any real-time images. 7. Designing caption.html such that once the model processes the uploaded image, the result must be displayed on the caption.html webpage along with that particular image. 8. Trying to display result on the same page. | | |
| **………………………………..**  **Signature of Internal Guide with date** |  | **………………………………..**  **Signature of Chairperson**  **with date** |