SOCIAL MEDIA SHORTCUT WORD TO PROPER ENGLISH WORD

Submitted by

Suharto Das

ID : 171-35-1980

Submitted to

Md Sanzidul Islam

Lecturer

Department of Software Engineering

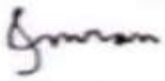
Daffodil International University

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science and Engineering in Software Engineering.

APPROVAL

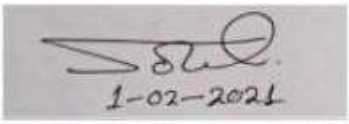
This Thesis titled " social media shortcut word to proper English word' .submitted by Suharto das ,171-35-1980 to Department of Software Engineering ,Daffodil international University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Software Engineering and approved as to its style and contents

BOARD OF EXAMINERS



Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University



Internal Examiner 1

Md. Khaled Sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

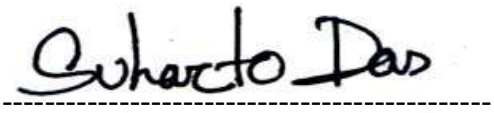
Md. Maruf Hassan
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Prof. Dr. Mohammod Abul Kaashem
Professor
Department of Computer Science and Engineering
Dhaka University of Engineering and Technology, Gazipur

DECLARATION

It is hereby declare that this thesis has been done by **Suharto Das** under the supervision of **Md Sanzidul Islam, Lecturer, Departments of Software Engineering, Daffodil International University.** It is also declare that neither this thesis nor any part of this has been submitted elsewhere for award of any degree.



Name- Suharto Das

ID no- 171-35-1980

Batch- 22

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

Certified by:

Md Sanzidul Islam

Lecturer

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

ACKNOWLEDGMENT

First of all, I am grateful to **The Almighty God** for giving me the ability to complete this research.

Today I am felling proud for myself. Because, to be a student of **Daffodil International University**. And I am thankful to Daffodil International University for giving me a chance to prove myself by showing this research . I would like to express my heartiest thanks to **Dr. Imran Mahmud** (Associate Professor and Head).

I express my sincere thanks to my respected my Supervisor **Md Sanzidul Islam**, who guided me through the research also gave valuable suggestions and guidance for completing the research. He helped me to understand the intricate issues involved in research - making besides effectively presenting it. These intricacies would have been lost otherwise. My research has been a success only because of his guidance.

At last I also thank my parents for their unceasing encouragement and support. I also place on record, my sense of gratitude to one and all who, directly or indirectly, have lent their helping hand in this research. And also to my friends who supported me in with basic concepts for this research.

Table of Contents

List of Figure

ABSTRACT

Day by Day, the trend of using social media has increased among the people and with the increasing rate the use of Banglish (Merge of Bangla and English) and shortcut words has also increased. In this research, Banglish and Shortcut words have been fully converted to English.

For this type of conversation, we have used the CNNs method and the method consist of multiple layer and these layers are connected to each other. This method is considered to be the best method because it does not need any feature extraction. The convolutional neuron network is used in many areas such as image and pattern recognition, speech recognition, natural language processing and video analysis. In this research we have used CNN method because we have decided to use computer vision as some of the words are close to each other and at first we need to convert all the shortcut words into images for pre-processing the data.

With the help of CNN method searched for the Banglish and shortcut words that people use for their daily conversation. At first we find the different representative of a single word and then we converted those shortcut and Banglish words to main word using CNNs method.

CHAPTER-1

INTRODUCTION

1.1 Background:

Now a day's the use of acronyms and abbreviations on social media are used vastly around the world. Mostly teenagers and young adults use simplified languages in their conversations. Like every other country Bangladesh is one of them and mostly everyone use this coded language in their conversation.

These shortcut languages are mostly used for text messaging on Facebook, Instagram, WhatsApp and on many other apps. These conversations are normally in Bengali but in English transcript which is very commonly known as (Banglish). In the last 10 years the use and the number of people use shortcut language have increased. It is very easy and a convenient way of communication as all educational group of people can use it and not only that even online sellers find it easy to communicate with buyers.

Few examples of shortcut languages are ttyl(talk to you later), omg(oh my god), idk(I don't know),plm(Problem),slv (Slove) etc. These shortcut conversations are used to save time and some phones have limited number of character per text message, some mobile phones have additional features of these shortcut words. Not every mobile phone or laptop/computer have the feature to write messages in Bengali, they needs to be installed separately.

1.2 Motivation of Thesis:

During the period of global pandemic every country has been using online/social media platform for teaching, so the use of shortcut languages have increased. One day while working on my project I contacted one of my friends regarding some materials. Our conversation involved text messaging in Banglish and shortcut words. I am familiar with both of the languages but I suddenly discovered that his way of using words are different than mine and other fellow classmates/friends for example some people write someone as (sum1, before as b4 etc.) I found it very difficult as his words were even shorter than others normally use. I then came up with the idea of creating a platform where you can use the shortcut words but it will automatically translate it into simple English, which will be easier for people to understand.

After researching a little further I saw that many countries use this shortcut languages and they include abbreviation, acronyms, slangs and less use of common grammar. Even in Bangladesh people do the same, so in order to save the languages from distortion I am going to use this software. It will help us to improve the understanding of the languages and also we can easily communicate with a friend or colleague (mainly in foreign countries) without bending the languages and it will help them and us to understand each other better.

1.3 Problem Statement:

These shortcut languages trends has been passed on to school students which can have effect on their academic. The shortcut words are also being used in our daily life for example (LOL, OMG, BRB) which is making us illiterate. The use of shortcut words are making people develop bad grammar and making the society incapable of structuring logical sentences. This shortcut social media words promoting bad grammar as people have adapted themselves to writing slang that even if they want to write proper but they cannot. The problem with using Banglish is that it will make people forget to write proper and clear Bengali as most of the schools use online platform as educating school/college students. This is not only a problem for our country but it's happening globally, some schools have even reported that the students used the shortcut languages in their exam paper.

1.4 Research Questions:

Question 1 : How we can create main word from the shortcut words?

Question 2 : How will this research impact on the human and the human society?

1.5 Objective:

While connecting with others through these social media's, most of the people use shortcut words in order to reduce more efforts to type full words. Using any word or sentence in shortcut can be a distortion of the word. Again, some people cannot understand the use or proper pronunciation of the word or sentence and sometimes it can be a great hazard for the people. Moreover, it is also noticed that some people also forgets the original spelling of the sentence that has been typed shortcut.

So, to solve this problems, I think the tool I intend to make is a time reliable tool. It will also solve the problem of word distortion. While typing a word in shortcut, my tool will suggest the proper and main word of the shortcut word. It will also let people remind the original spelling of the word just by clicking and selecting the proper word for the shortcut word. Thus, this way people will not be confused and gets in hazard.

1.6 Limitations:

We didn't get much data. Google didn't get the kind of data we wanted. So we had to create a dataset. Since we had created dataset, the amount of data in the dataset is somewhat less. In addition, different people use different types of shortcut words. So collecting shortcut words for all words is a big challenge.

CHAPTER-2

LITERATURE REVIEW

This paper tells that CNN model are similar to ordinary neural network, they have hidden layers containing neurons. After the test were made accurately for CNN-softmax and CNN-svm, the results were almost the same. If the base of the CNN model is refined the results may be improved. By Abien Fred M.Agarap et al [1]. The paper talks about an advanced word predicting system (Fast Type) which excels typical limitation of standard techniques. Fast Type is a combined statistical and rule based methods developed to improve Keystroke saving. It is used to cut the number of keystroke use, which helps people with special needs and native users to reduce misspelling and the non-local people to use correct words. By Carlo Aliprandi et al [4]. In this paper we discuss how features relying networks are employed for English article correction, instead of relying on human skills. The network is known as convolutional neural network. This network can be used on both error annotated and non-annotated corpus by Chengjie sun et al [6]. The purpose of the hybrid model for word prediction problem is to recommend next word in a sentence. This problem can be solved by two methods, syntactic or semantic analysis. This two features cannot be used individually as it will not give any practical result. On the contrary if these two methods are used together it can provide state-of-the-art-result.{88} In order to predict the method of characterization of social relation from word-based contents unsupervised learning algorithm meta-data is used. It has been found that in order to improve the prediction accuracy a limited number of information needs to be presented. By Georg Sun et al [7]. Federated Algorithm is a network language which allows its client (mobile phone) to better prediction recall by giving training on a high quality data set. The work shows the possibility of the training without transferring its client's data to the server and also gives control over the use of their data by Andrew

et al [9]. There has been many problem in classifying short texts because of lack of contextual information. It creates difficulties for the deep neural network to produce similar binary codes for semantically related texts. This problem was overcome by removing nouns and used the verbs to improve the short text classification. When convolutional neural networks have been a great success in single-label image classification but still it failed to clearly label dependencies in an image. So in order to address to this problem CNN-RNN framework has been combined together.by Jiang Wang et al [12]. This paper demonstrates how modern CNN-RNN hybrid architecture model used to recognize hand written scripts of Hindi and Bangla. In order to recognize the problems for languages where suitable size of data is not available, pre-training the network on synthetic data can resolve the problem. By kartik Dutta et al [13]. The convolutional neural network acknowledges automatic detection and acknowledgement of text in natural images. The network is trained to perform word recognition on the whole proposal region. In order to detect fast region proposal methods to perform word detection the channel uses a novel combination of complementary proposal generation techniques. After examining the steps it is finally ready to be used in the real world application of text spotting, which will allow hours long news footage to be instantly searchable through a text query. Max Jaderberg et al [18]. The purpose of this paper is to build a language model to predict a correct word in a sentences. It will not only save time but will help the disable people and will reduce spelling mistakes by Md.Masudul Haque et al [19]. A novel convolutional architecture named as genCNN which uses as a word sequence prediction. It is different from the neural network and the existing feedforward networks. This new model is fast, easy to train and after the experiments made on text generation and n-best re-ranking in machine translation it can significantly improve upon state of arts by MingxuanWange et al [20].

In this paper hybrid machine translation has been used to translate Chinese-to-English and English-to-Chinese by using BJTU-NLP system. To further improve the result we have taken data from Wikipedia database and pre-processing and post-processing rules are applied to improve performance.

CHAPTER-3

METHODOLOGY

Most important and vital part of any research work is data collection and data pre-processing. And as we work with a unique data set we had to spend most of our time in data set collection. We collected generally shortcut texting words for that we have to find out the different representation of a single word. We used 71 words and for training we used 10-15 different representations of that word and 7-10 different representations for testing the model. In figure 3.1 we produce an example of our data.

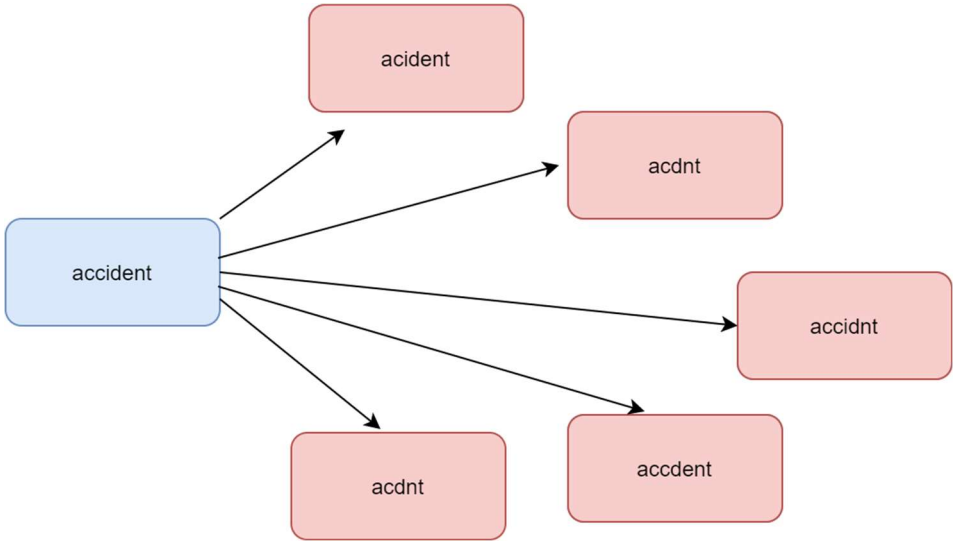


Figure 3.1: Dataset Example

In figure 3.2 we described our workflow to complete the whole process.

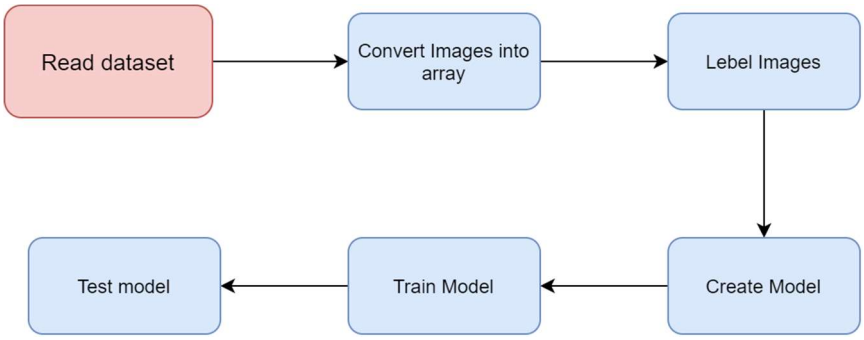


Figure 3.2: Proposed workflow

3.1 Data pre-processing

We can't create any rule or find any pattern to labeling the shortcut typing and actual words. Suppose we write "hlw" instead of hello in vector distance count hlw is close to hi so from that perspective "hlw" will be "hi". So we decided to use computer vision. So that we can work from the viewing perspective for that we need to convert all the shortcut words into images. In figure 3.2 we described how we converted plain words into images.

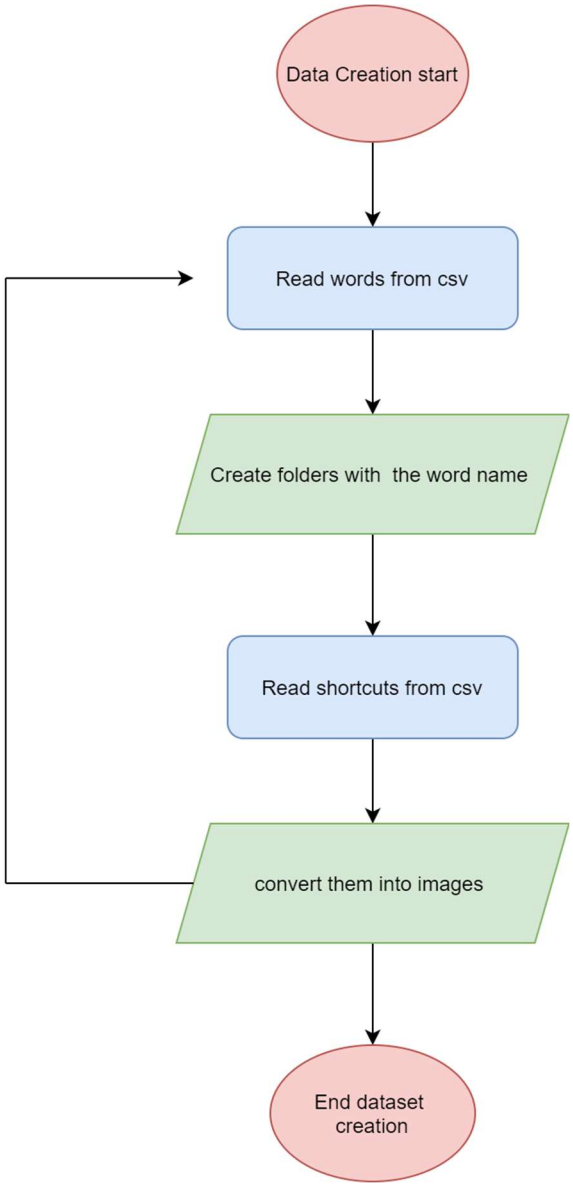


Figure 3.3: plain words to images

3.2 Labeling dataset

As we had 71 unique words and the shortcut with that we had to label them with their actual data words. Cause we will use supervised learning. Data labeling is important as it is the base for any AI project. This named information goes about as an informational collection which is taken care of in calculations to prepare different AI models.

AI models generally need loads of information marking for each task which is called preparing information and this named information ought to be of high and exact quality for a Machine Learning model to work precisely in a true situation. This Labeled information assists AI with perceiving different articles, shapes and examples. In this case we used the folder name as labeling words for the shortcuts. In figure 3.3 3rd step we will find our labeling and in figure 3.4 We described how we read the images from the dataset.

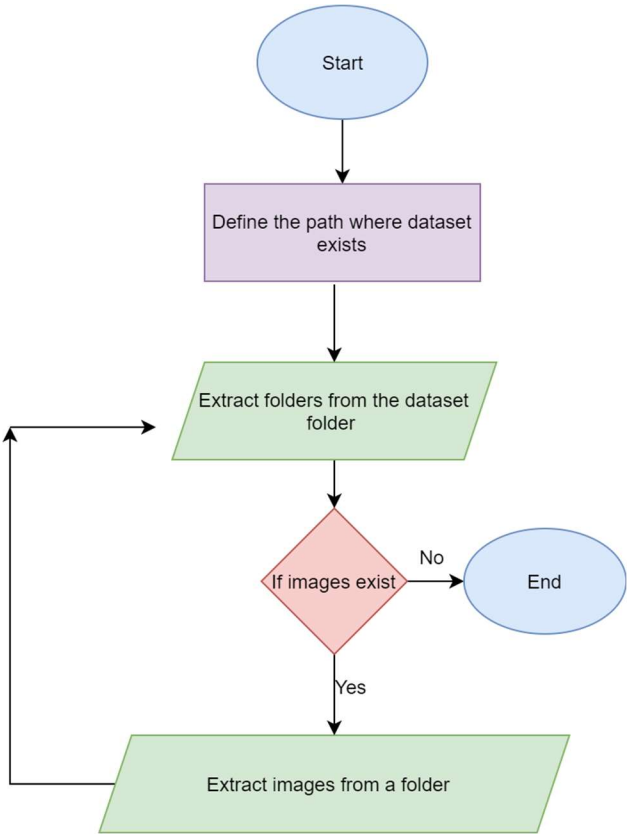


Figure 3.4: Image reading from dataset

3.3 Label Encoding:

We had 71 unique words but we couldn't represent them as strings. We had to represent them into numeric values. For that we used label encoding. Label Encoding alludes to changing over the names into the numeric structure to change over it into the machine-lucid structure. AI calculations would then be able to choose in a superior manner on how those marks should be worked. It is a significant pre-preparing venture for the organized dataset in regulated learning.

3.4 Feature extraction:

Feature extraction includes diminishing the number of assets needed to portray a huge arrangement of information. When performing the examination of complex information one of the serious issues originates from the number of factors included. Examination with countless factors, for the most part, requires a lot of memory and calculation power, likewise, it might make an order calculation overfit to preparing tests and sum up ineffectively to new examples. Highlight extraction is an overall term for techniques for developing blends of the factors to get around these issues while depicting the information with adequate precision. Many AI specialists accept that appropriately enhanced element extraction is the way to compelling model development. In figure 3.5 we described our feature extraction technique.

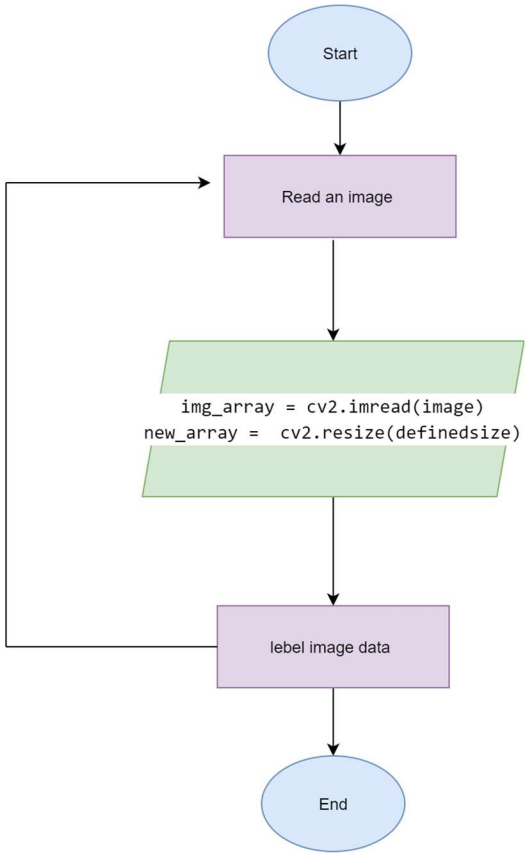


Figure 3.5: Feature extraction technique.

In below figure 3.6 we visualize an example of our extracted feature.

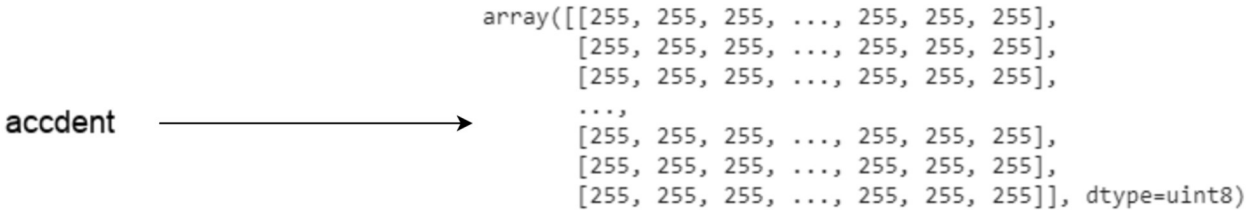


Figure 3.6: Images to array conversion

3.5 Proposed Model:

We mainly used convolutional neural networks. CNNs are regularized forms of multilayer perceptrons. Multilayer perceptrons generally mean completely associated networks, that is, every neuron in one layer is associated with all neurons in the following layer. The "completely connectedness" of these organizations makes them inclined to overfitting information. Common methods of regularization incorporate adding some type of extent estimation of loads to the misfortune work. CNNs adopt an alternate strategy towards regularization: they exploit the progressive example in information and collect more mind boggling designs utilizing more modest and easier examples. Along these lines, on the size of connectedness and intricacy, CNNs are on the lower extraordinary.

Convolutional networks were propelled by natural cycles in that the availability design between neurons takes after the association of the creature visual cortex. Individual cortical neurons react to upgrades just in a confined district of the visual field known as the open field. The responsive fields of various neurons somewhat cover with the end goal that they cover the whole visual field.

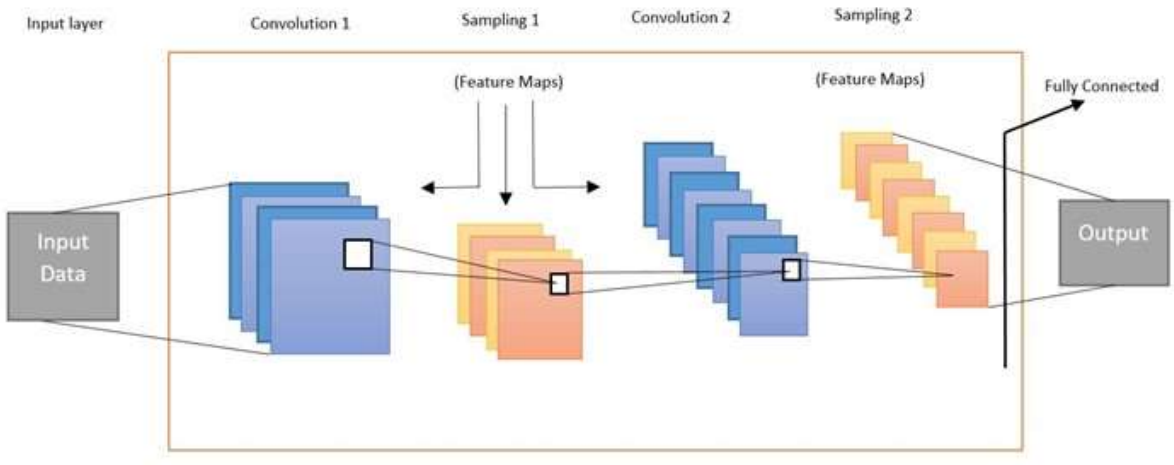


Figure 3.7: Proposed model

In the figure below we described the actual summary of our model.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 98, 98, 32)        320
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 49, 49, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 47, 47, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 23, 23, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 33856)             0
_________________________________________________________________
dense (Dense)                (None, 71)                2403847
=================================================================
Total params: 2,422,663
Trainable params: 2,422,663
Non-trainable params: 0
_________________________________________________________________
```

Figure 3.8: Summary of the model

And the visual representation of our model architecture is given in figure 3.8

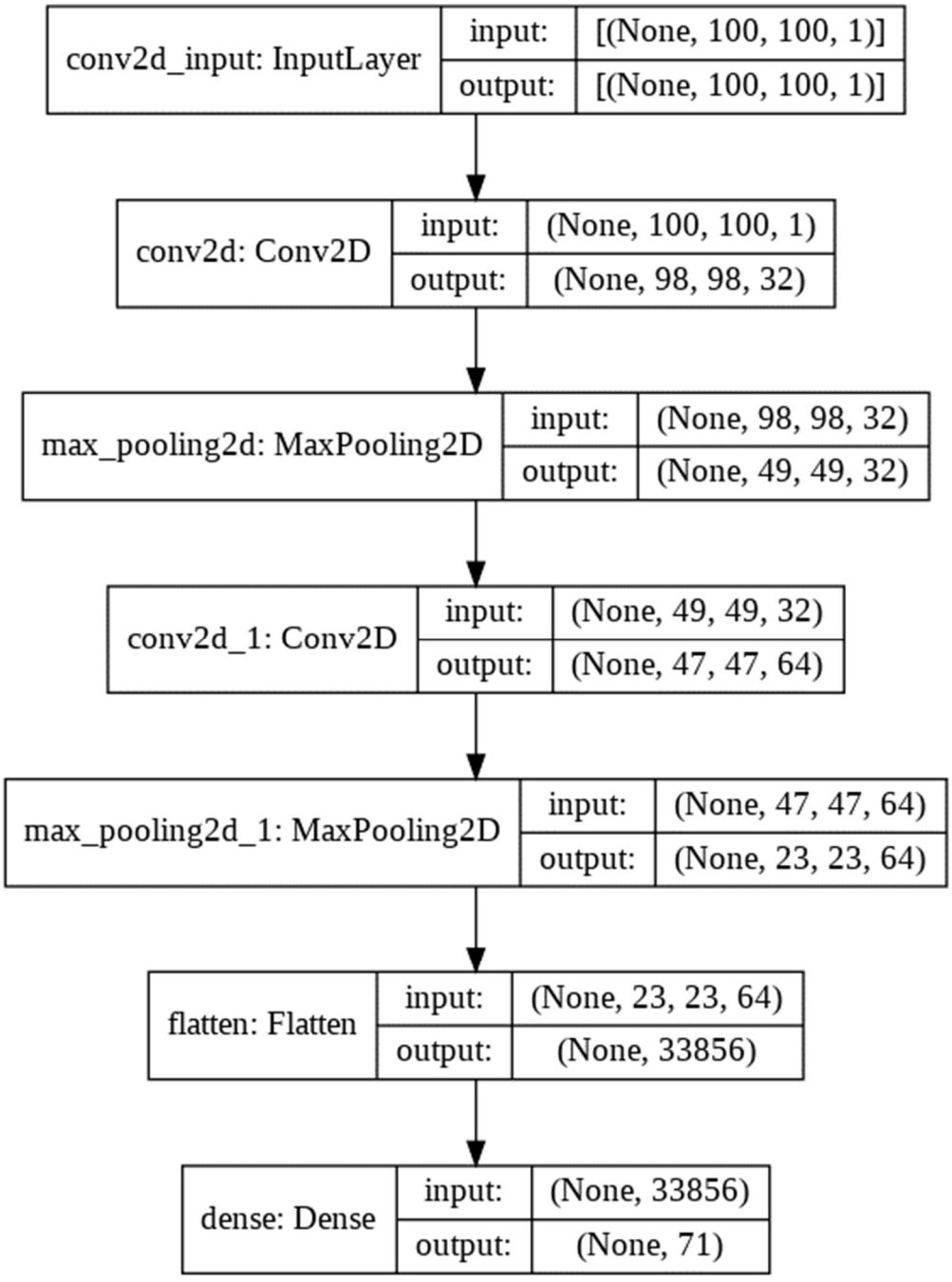


Figure 3.9: Model architecture.

CHAPTER-4

RESULTS AND DISUSSION

4.1 Precision:

Precision actually monitors the accuracy of the model when it's predicting the true values. Below we provide the formula of finding precision.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

4.2 Recall:

Recall actually monitors the accuracy of the model when it's predicting the false values. Below we provide the formula of finding recall.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

F1-score:

This is actually a statistical analysis of the performance of a model. Which actually combined precision and recall.

Below we provided the formula of finding f1-scores.

$$F_1\text{−score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

For better and clear understanding of classification we plotted a CM (confusion matrix) as we had a total 71 class. It's a large confusion matrix but understandable. In figure 4.1

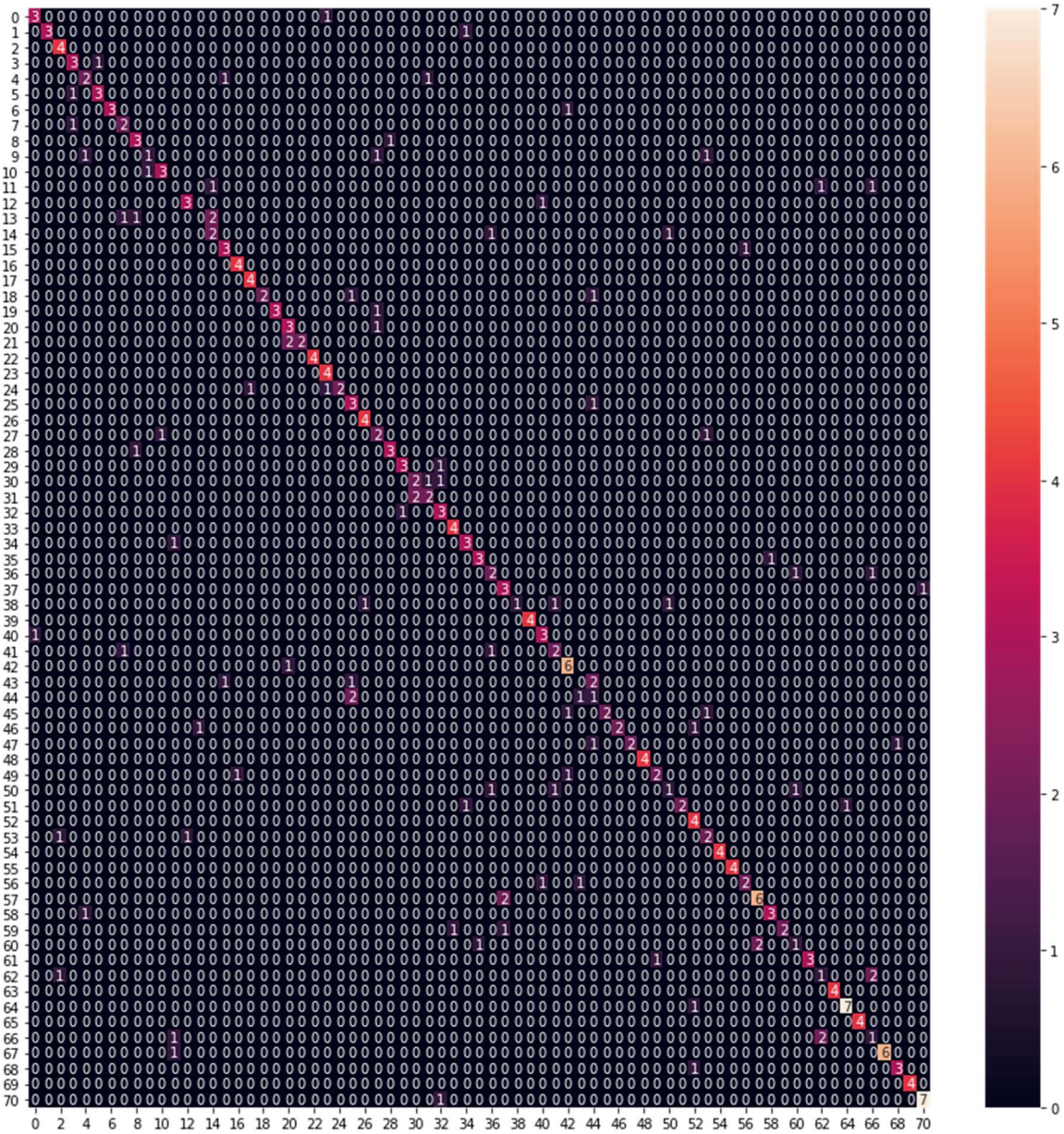


Figure 4.1: Confusion matrix of the model.

4.3 Training accuracy:

Above we described validation accuracy and the perfectness of the model with unseen dataset. Now we will describe how our model works on trained images.

In figure 4.2 we showed the accuracy of the model on the training dataset.

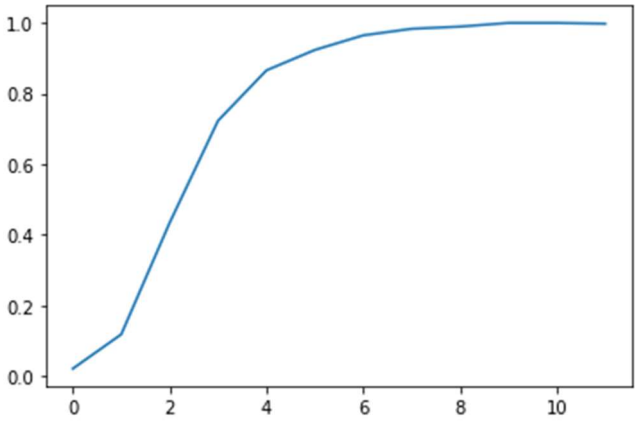


Figure 4.2: Training accuracy

If we look at the graph we can see the increase of accuracy with epochs. We used 12 epochs where we can avoid overfitting and at the same time we can achieve the highest accuracy.

In figure 4.3 we can see the decrease of loss with the epochs. This accuracy and loss is vise-versa with less loss we will get better accuracy.

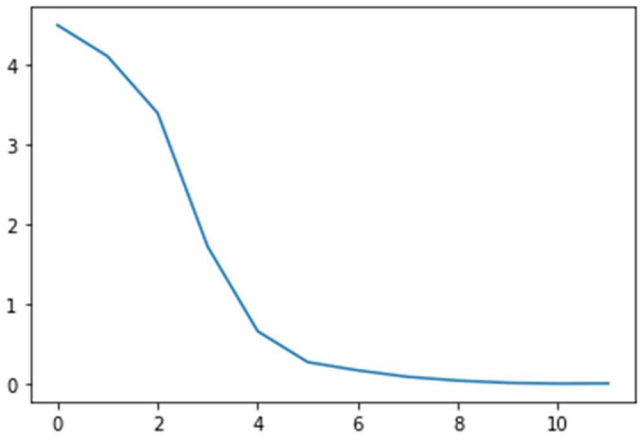


Figure 4.3: Testing accuracy

In this work we used convolutional neural networks for classifying word's images.

Below we described the classification report. Where we find out the precision, recall, f1-score and finally the accuracy for the unseen dataset to the model and we got 75% accuracy in validation.

	precision	recall	f1-score	support
0	1.00	0.75	0.86	4
1	1.00	0.75	0.86	4
2	0.67	0.50	0.57	4
3	0.60	0.75	0.67	4
4	0.75	0.75	0.75	4
5	0.75	0.75	0.75	4
6	1.00	1.00	1.00	4
7	0.50	0.67	0.57	3
8	0.60	0.75	0.67	4
9	0.50	0.25	0.33	4
10	0.50	0.75	0.60	4

11	0.00	0.00	0.00	3
12	1.00	1.00	1.00	4
13	0.00	0.00	0.00	4
14	0.60	0.75	0.67	4
15	0.67	0.50	0.57	4
16	0.67	1.00	0.80	4
17	0.80	1.00	0.89	4
18	1.00	0.75	0.86	4
19	1.00	0.75	0.86	4
20	0.50	1.00	0.67	4
21	1.00	0.50	0.67	4
22	1.00	1.00	1.00	4
23	1.00	1.00	1.00	4
24	1.00	1.00	1.00	4
25	0.50	0.75	0.60	4
26	0.80	1.00	0.89	4
27	0.75	0.75	0.75	4
28	0.75	0.75	0.75	4
29	1.00	0.75	0.86	4
30	0.67	0.50	0.57	4
31	0.75	0.75	0.75	4
32	0.67	1.00	0.80	4
33	0.80	1.00	0.89	4
34	0.80	1.00	0.89	4
35	1.00	0.75	0.86	4
36	0.60	0.75	0.67	4
37	0.75	0.75	0.75	4
38	1.00	0.25	0.40	4
39	1.00	1.00	1.00	4
40	1.00	1.00	1.00	4
41	0.50	0.50	0.50	4
42	1.00	0.86	0.92	7

43	0.25	0.25	0.25	4
44	0.25	0.25	0.25	4
45	1.00	0.50	0.67	4
46	0.75	0.75	0.75	4
47	0.75	0.75	0.75	4
48	1.00	1.00	1.00	4
49	1.00	0.75	0.86	4
50	0.33	0.25	0.29	4
51	1.00	0.50	0.67	4
52	0.67	1.00	0.80	4
53	0.67	0.50	0.57	4
54	1.00	1.00	1.00	4
55	1.00	1.00	1.00	4
56	0.67	0.50	0.57	4
57	0.70	0.88	0.78	8
58	0.60	0.75	0.67	4
59	1.00	0.50	0.67	4
60	1.00	0.25	0.40	4
61	1.00	1.00	1.00	4
62	0.44	1.00	0.62	4
63	0.80	1.00	0.89	4
64	0.88	0.88	0.88	8
65	0.80	1.00	0.89	4
66	0.50	0.50	0.50	4
67	1.00	1.00	1.00	7
68	0.75	0.75	0.75	4
69	0.80	1.00	0.89	4
70	0.88	0.88	0.88	8
accuracy			0.75	300
macro avg	0.76	0.74	0.73	300
weighted avg	0.77	0.75	0.74	300

CHAPTER-5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Findings and Contributions

In this study, we find out while connecting with others through these social media's, most of the people use shortcut words in order to reduce more efforts to type full words. Using any word or sentence in shortcut can be a distortion of the word. Again, some people cannot understand the use or proper pronunciation of the word or sentence and sometimes it can be a great hazard for the people. Moreover, it is also noticed that some people also forgets the original spelling of the sentence that has been typed shortcut.

So, to solve this problems, I think the tool I intend to make is a time reliable tool. It will also solve the problem of word distortion. While typing a word in shortcut, my tool will suggest the proper and main word of the shortcut word. It will also let people remind the original spelling of the word just by clicking and selecting the proper word for the shortcut word. Thus, this way people will not be confused and gets in hazard.

5.2 Recommendations for Future Works

We have not a huge dataset. So dataset is the one of the limitation of the research. For solving shortcut word to main word we need huge dataset which do not have currently. In future we increase this dataset and also add **BEM** with **ShcWd**(Shortcut word) .

REFERENCES

[1] Abien Fred M. Agarap. (2019, February 7). An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification.

[2] Alice E. Neels, N. K. (2015). TEXT CORRECTION PROCESSING.

[3] Bluche, T., H. N., & C. K. (2013, August). Feature extraction with convolutional neural networks for handwritten word recognition.

[4] Carlo Aliprandi, Nicola Carmignani, Nedjma Deha, Paolo Mancarella, & Michele Rubino. (2008, February). Advances in NLP applied to Word Prediction.

[5] Ce Zhang, Xin Pan, Huapeng, Andy Gardiner, Isabel Sargent, Jonathon Hare, & Peter M. Atkinson. (2018, June). A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification.

[6] Chengjie Sun, Xiaoqiang Jin, Lei Lin, Xiaolong Wang, & Yuming Zhao. (2015). Convolutional Neural Networks for Correcting English Article Errors. pp. LNAI 9362, pp. 102–110, 2015.

[7] Georg Groh, & Jan Hauffa. (2011). Characterizing Social Relations Via NLP-based Sentiment Analysis. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.*

[8] Haddad, B. (2007, December). Detection and Correction of Non-Words in Arabic: A Hybrid Approach.

[9] Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Sean Augenstein, F. B., Eichner, H., . . . Ramage, D. (2019). FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION.

[10] Jason Tyler Griffin, Jerome Pasquero, & Donald Somerset McKenzie. (2015, Sep 1). IN-LETTERWORD PREDICTION FOR VIRTUAL KEYBOARD.

[11] Jason Tyler Griffin,, Alistair Robert Hamilton,, & Donald Somerset McCulloch. (2015). TOUCHSCREEN KEYBOARD WITH CORRECTIVE WORD PREDICTION.

[12] Jiang Wang, Yi Yang, Junhua Mao, & Zhiheng Huang. (2016). CNN-RNN: A Unified Framework for Multi-label Image Classification. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[13] Kartik Dutta, Praveen Krishnan, Minesh Mathew, & Jawahar. (2018, April). Towards Accurate Handwritten Word Recognition for Hindi and Bangla.

[14] KUKICH, K. (1992). Techniques for Automatically Correcting Words in Text. *ACM Computing Surveys*, (pp. Vol. 24, No. 4).

[15] Kushwah, K. K. (2019). Hindi word correction using micro-parsing. *Int. J. Arts and Technology*, Vol. 11, No. 1.

[16] Lee, J., & S. S. (2006). Automatic Grammar Correction for Second-Language Learners. *INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA.*

[17] Manuel Carreiras, Blair C. Armstrong, Manuel Perea, & Ram Frost. (2014, February). The what, when, where, and how of visual word recognition.

[18] Max Jaderberg, Karen Simonyan, & Andrea Vedaldi. (2014, Dec 4). Reading Text in the Wild with Convolutional Neural Networks.

[19] Md. Masudul Haque, Md. Tarek Habib, & Md. Mokhlesur Rahman. (November 2015). AUTOMATED WORD PREDICTION IN BANGLA LANGUAGE USING STOCHASTIC LANGUAGE MODELS. *International Journal in Foundations of Computer Science & Technology*, Vol.5, No.6,.

[20] MingxuanWang, Zhengdong Lu, Hang Li, Wenbin Jiang, & Qun Liu. (2015, April 24). A Convolutional Architecture forWord Sequence Prediction.

[21] Simard, P. Y., D. S., & J. C. (6 August 2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. *7th International Conference on Document Analysis and Recognition (ICDAR 2003)*, (pp. 2-Volume Set, 3). Edinburgh, Scotland, UK.

[22] Théodore Bluche, Hermann Ney, & Christopher Kermorvant. (n.d.). TANDEM HMM WITH CONVOLUTIONAL NEURAL NETWORK FOR HANDWRITTEN WORD RECOGNITION. *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference* .

[22] Wang, D., X. Y., J. X., & Y. C. (2015). A Hybrid Transliteration Model for Chinese/English Named Entities —BJTU-NLP Report for the 5th Named Entities Workshop. *Proceedings of the Fifth Named Entity Workshop joint with 53rd ACL and the 7th IJCNLP*, (pp. 67–71).

[23] X, H., Goulart,, & D. L., M. (2018). Hybrid Model For Word Prediction Using Naive.

[24] Xiaohang Ren, Kai Chen , & Jun Sun. (2016, april 7). A CNN Based Scene Chinese Text Recognition Algorithm With Synthetic Data Engine.

[25] Yuan, A., G. B., P. Y., Y. G., & Xinting. (0212). Handwritten EnglishWord Recognition based on Convolutional Neural Networks. *2012 International Conference on Frontiers in Handwriting Recognition*, (pp. 978-0-7695-4774).

[26] Yunchao Wei, Wei Xia,, Min Lin, Junshi Huang, & Bingbing Ni, Jian Dong. (n.d.). HCP: A Flexible CNN Framework for Multi-Label Image Classification.

[27] Zhan, J., & Dahal, B. (2017). Using deep learning for short text.

Appendix

List of Abbreviation

BEM	Bangla sentence in English format (Ami ekta problem slv korlam.)
ShcWd	Shortcuts Word
CM	Confusion Matrix
CNN	Convolutional Neural Network

Percentage of Plagiarism

Turnitin Originality Report

Processed on: 02-Feb-2021 14:05 +06
ID: 1499780086
Word Count: 5048
Submitted: 1

171-35-1980 By Suharto Das

Similarity Index	Similarity by Source
19%	Internet Sources: 12% Publications: 8% Student Papers: 15%

3% match (student papers from 30-Oct-2017) Submitted to United International University on 2017-10-30
3% match (student papers from 15-Aug-2020) Submitted to British University in Egypt on 2020-08-15
2% match (publications) R. Sathishkumar, K. Kalaiarasan, A. Prabhakaran, M. Aravind, "Detection of Lung Cancer using SVM Classifier and KNN Algorithm", 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019
1% match (student papers from 15-Sep-2020) Submitted to Manukau Institute of Technology on 2020-09-15
1% match (publications) Max Jaderberg, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, "Reading Text in the Wild with Convolutional Neural Networks", International Journal of Computer Vision, 2015
1% match (student papers from 30-Apr-2016) Submitted to Institute of Technology, Nirma University on 2016-04-30
1% match (Internet from 30-Sep-2020) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4316/151-15-364.151-15-372.pdf?isAllowed=y&sequence=1
1% match (student papers from 26-Jul-2020) Submitted to Universiti Teknologi MARA on 2020-07-26
1% match (Internet from 29-Mar-2020) https://www.scribd.com/document/402407862/Scipy-Lecture-Notes-pdf
1% match (Internet from 01-Apr-2020) https://www.slideshare.net/RaihanMahmud5/remote-doctor-project-report
1% match () http://usir.salford.ac.uk/id/eprint/42351/1/%4000344526_Mohd%20Nadhir%20Ab%20Wahab_Thesis.pdf
1% match (student papers from 09-Apr-2018) Class: Article 2018 Assignment: Journal Article Paper ID: 943487689
< 1% match (Internet from 22-Apr-2016) http://aclweb.org/anthology/W/W15/W15-39.pdf
< 1% match (Internet from 19-Nov-2020) https://medium.com/swlh/machine-learning-algorithms-with-scikit-learn-ec6181365ba