

MALIGANT COMMENT CLASSIFIER PROJECT

Submitted by:

Mandar M Joshi

**ACKNOWLEDGMENT**

For every NLP project it is very essential and necessary to understand the basic idea about creating a model. Internet and open source always helps to understand thoroughly to enhance basic understanding of the business model.

**INTRODUCTION**

* Business Problem Framing

There has been a remarkable increase in the cases of cyberbullying and trolls on various social media platforms. Many celebrities and influences are facing backlashes from people and have to come across hateful and offensive comments. This can take a toll on anyone and affect them mentally leading to depression, mental illness, self-hatred and suicidal thoughts.so in order to detect these messages need to build a machine to identify it.

* Conceptual Background of the Domain Problem

The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. Although researchers have found that hate is a problem across multiple platforms, there is a lack of models for online hate detection.

* Review of Literature

Online hate, described as abusive language, aggression, cyberbullying, hatefulness and many others has been identified as a major threat on online social media platforms. Social media platforms are the most prominent grounds for such toxic behaviour

Internet comments are bastions of hatred and vitriol. While online anonymity has provided a new outlet for aggression and hate speech, machine learning can be used to fight it. The problem we sought to solve was the tagging of internet comments that are aggressive towards other users. This means that insults to third parties such as celebrities will be tagged as unoffensive, but “u are an idiot” is clearly offensive.

* Motivation for the Problem Undertaken

Our goal is to build a prototype of online hate and abuse comment classifier which can used to classify hate and offensive comments so that it can be controlled and restricted from spreading hatred and cyberbullying.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

* Data Sources and their formats

What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

* Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

* Data Inputs- Logic- Output Relationships

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

* State the set of assumptions (if any) related to the problem under consideration

Here, you can describe any presumptions taken by you.

* Hardware and Software Requirements and Tools Used

Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

The data set contains the training set, which has approximately 1,59,000 samples and the test set which contains nearly 1,53,000 samples. All the data samples contain 8 fields which includes ‘Id’, ‘Comments’, ‘Malignant’, ‘Highly malignant’, ‘Rude’, ‘Threat’, ‘Abuse’ and ‘Loathe’.

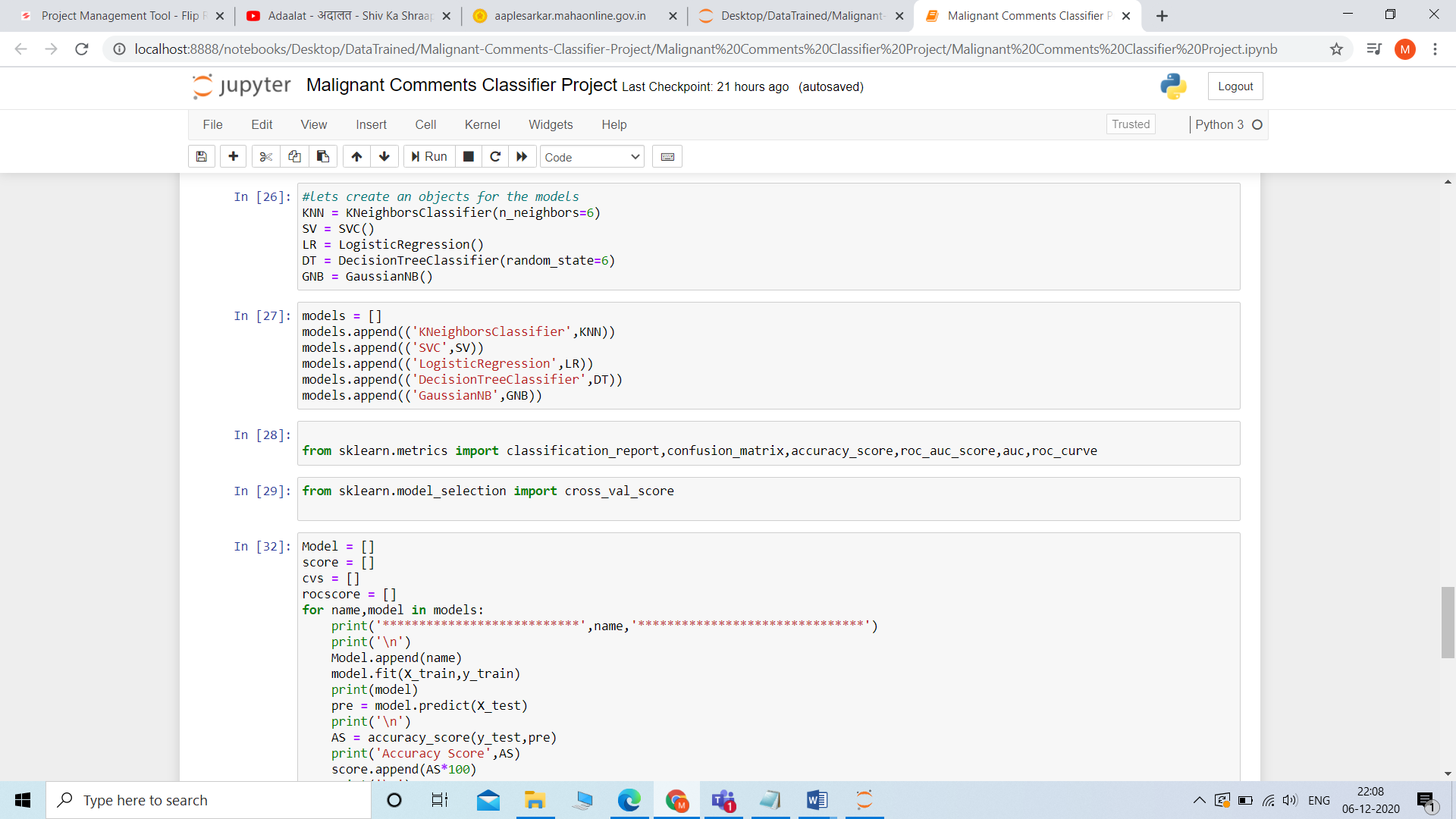
In which our goal is to focus on”Loathe” column and analyse the text for in “comments” column.

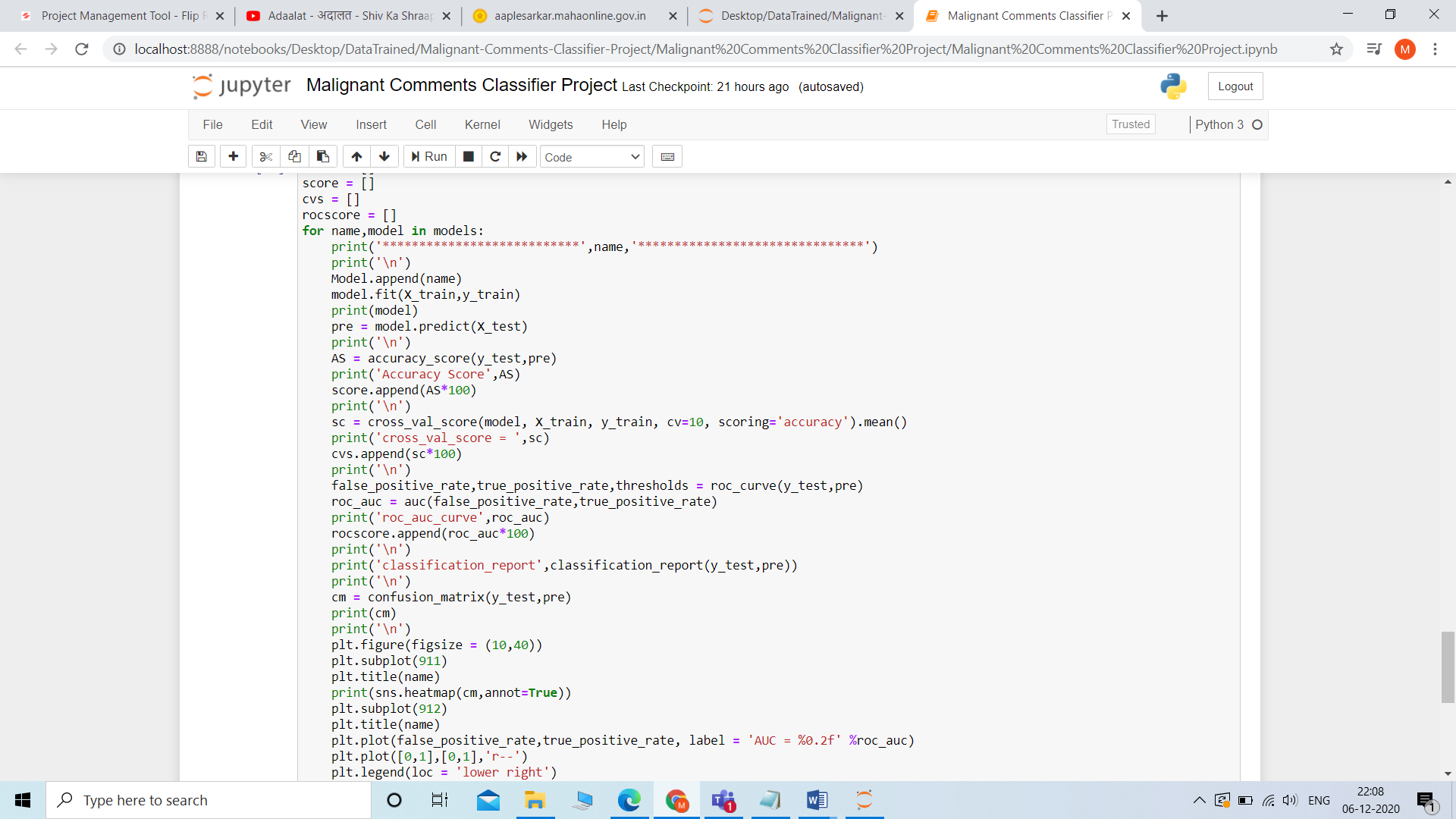
* Testing of Identified Approaches (Algorithms)

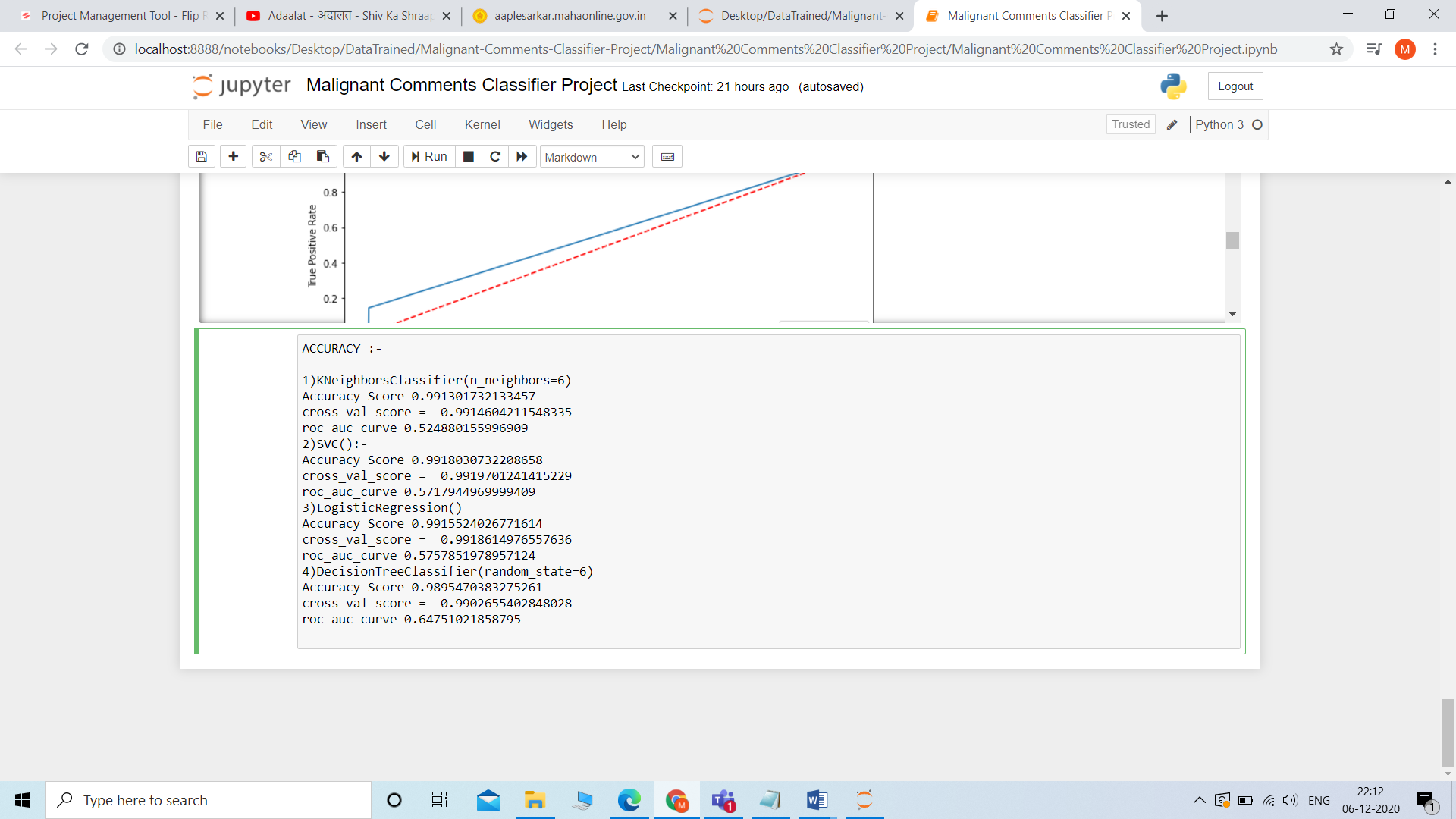
Following are the alogorithms that I use for this problem:-kneighboursclassifier,SVC,Logistic Regression,Decision Tree Classifier,GaussionNB.By selecting multiple algorithm we can actually able to see that which perform very well on the dataset.

* Run and Evaluate selected models

Following are the alogorithms that I use for this problem:-kneighboursclassifier,SVC,Logistic Regression,Decision Tree Classifier,GaussionNB.







* Key Metrics for success in solving problem under consideration

Confusion matrics used to see the accuracy score,also used cross\_val score and roc\_auc\_curve to validate the model effectively.

* Visualizations

Roc\_auc curve that is area under the curve gives the proportion of true positive rate verses false positive and the visualisation of confusion matrics to check the model.

* Interpretation of the Results

After doing preprocessing which incudes stopwords,word lemmatization technique,tfidf and count vectorization the dataset finally supplied to the different models to analyse and finalise the final model which works better as compare to other.

**CONCLUSION**

* Key Findings and Conclusions of the Study

1)KNeighborsClassifier(n\_neighbors=6)

Accuracy Score 0.991301732133457

cross\_val\_score = 0.9914604211548335

roc\_auc\_curve 0.524880155996909

2)SVC():-

Accuracy Score 0.9918030732208658

cross\_val\_score = 0.9919701241415229

roc\_auc\_curve 0.5717944969999409

3)LogisticRegression()

Accuracy Score 0.9915524026771614

cross\_val\_score = 0.9918614976557636

roc\_auc\_curve 0.5757851978957124

4)DecisionTreeClassifier(random\_state=6)

Accuracy Score 0.9895470383275261

cross\_val\_score = 0.9902655402848028

roc\_auc\_curve 0.64751021858795.

# here we can use SVC model which gives higher accuracy such as 99.18%

* Learning Outcomes of the Study in respect of Data Science

As in the particular dataset contains about 1,59,000 rows are present the only challenge is for the training as it takes a very long time in my case for every iteration.so in order to achieve this.in initial period I just fit the data and predict to check the time for the first time.

* Limitations of this work and Scope for Future Work

I don’t see any limitations only apart from higher training time but it completely depend on the hardware that we are using ,if we upgrade our system with a nice configuration, the chances of lagging decreases and will produce nice results.

Future scope is really nice to have more number of rows or column with more complex text for further deeper and complicated solution.