**1.INTRODUCTION**

**1.1 Overview**

Most of the consumers read online reviews before they decide to purchase a Mobile phone from any e-commerce website .Online Mobile applications has revolutionised the way consumers purchase mobile phones online as these apps have all the information regarding any mobile phone at users finger tips. Amazon is one of the best mobile applications which is considered as a treasure trove of all mobile reviews, and their review system is accessible across all channels presenting reviews in an easy-to-use format. So,There should be a system which analyses thousands of reviews of unlocked mobile phones sold on Amazon.com to find insights with respect to reviews, ratings, price and their relationships.

**1.2 Purpose**

The objective of the project is to build a model that predict the helpfulness of the review based on the review text . This helps in knowing the potential of the review which inturn helps the consumers to choose a better product according to their needs

**2. Literature survey**

**2.1 Existing system**

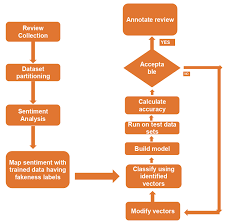
Amazon is one of the best mobile applications which is considered as a treasure trove of all mobile reviews, and their review system is accessible across all channels presenting reviews in an easy-to-use format. This platform contains numerous reviews about the products that are available in amazon.com but we may not know whether the review is useful is or not.

**2.2 Proposed system**

The objective of the system is to analyse the thousands of reviews of unlocked mobile phones sold at amazon.com to find insights with respect to ratings, reviews , price and their relationships.Corpus-based and knowledge-based methods can be used to determine the semantic similarity of review text. We will be using Natural language processing to analyse the sentiment ( positive or a negative) of the given review .

**3 THEORTICAL ANALYSIS**

**3.1 Block diagram**



**3.2 Hardware and Software designing**

HARDWARE:

1. Processor: Core i3 or above.

2. RAM: 4GB.

3. Other: Standard Monitor, Keyboard, Mouse etc.

SOFTWARE:

1. Programming Language: Python Programming.

2. Software Tool: Anaconda (Jupyter, Spyder).

3. Technologies: Deep Learning.

4. Framework: Flask

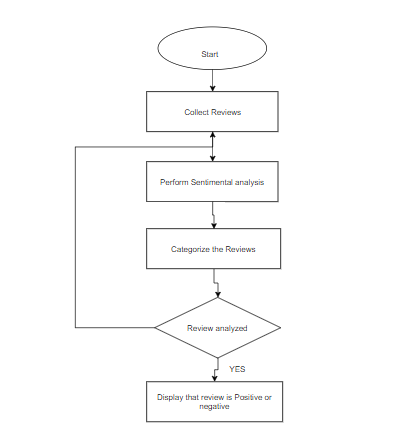
**4. EXPERIMENTAL INVESTIGATIONS**

Today, most e-commerce websites have a separate section where their customers can post reviews for products or service. Important information like customers' opinion on products, reasons for negative reviews, suggestions, etc., can be extracted from the posted reviews by performing sentiment analysis on them. Consumers can also assign a numerical value (i.e., rating) to the product or service they are reviewing. On Amazon.com the rating can be between 1 and 5 where 1 is the worst and 5 is the best. In some instances, there is a mismatch between a customer’s review and rating. It is important to identify the reviews with mismatched ratings since individual ratings are used to compute the average rating. Conventionally, a review text was converted to fixed-length feature vector using bag-of-words or bag-of-n-grams and these feature vectors were later used to train a shallow classifier such as naïve Bayes or support vector machine [1]. Although bag-of-words performed surprisingly well and was popular for many years, it has two major flaws. It loses ordering of words and doesn't consider the semantic relationship between words.

The data used in this study is a set of approximately 3.5 million product reviews collected from Amazon.com by Fang et al. Each review includes information on rating, product id, helpfulness, reviewer id, review title, review time, and review text. The rating is based on a 5-star scale.

In order to analyse the sentiment of Amazon.com reviews we built a model using recurrent neural networks (ANN) with gated recurrent unit (GRU) that learned low-dimensional vector representation of reviews using paragraph vectors and product embeddings. We first converted Amazon.com product reviews to fixed-length feature vectors using paragraph vectors. These feature vectors were then grouped by product and sorted in temporal order. Each group was used to train an ANN with GRU. The vectors generated in the penultimate layer of the RNN are called product embeddings. These embeddings capture important information like product qualities and temporal relations among reviews. We then concatenated product embeddings with fixed-length vectors generated by paragraph vectors and trained a support vector machine. We also developed a user interface to tackle the review-rating mismatches. There are situations where a user may write a highly positive review but give it 1 or 2 stars or write a highly negative review but give it 4 or 5 stars [13]. Although such cases are rare, they create confusion among users who read these reviews. To address this issue, we labeled reviews with rating 1 and 2 are as ‘negative’, reviews with rating 3 as ‘neutral’, and reviews with rating 4 and 5 as ‘positive’. We then developed a web service application that uses our classifier to predict a class from the user review. If the predicted class and the class that the submitted rating belongs to are different a warning is submitted to the user so that they can review and correct their rating.

**5 .FLOW CHART**



**6 .RESULT**

A model which can predict the helpfulness and the nature of the reviewsof various mobile phones that are present in amazon.com is developed.

**7. ADVANTAGES AND DISADVANTAGES**

**Advantages:**

1.Nature of the review can be predicted.

2.Helpfulness of a review can be known.

3.Fradulent reviews can be avoided.

**Disadvantages:**

1.The data sholud be properly evaluated otherwise wrong predictions may occur

2. The data should be properly trained

**8. APPLICATIONS**

### 1. Appbot: App review and rating analysis for mobile phones

2. Apps

3. Review based buying and selling

**9 .CONCLUSION**

This system analyses thousands of reviews of mobile phones that are sold on amazon.com .It can predict whether a review is genuine or not and it can also predict the helpfulness of a particular review.

**10 .FUTURE SCOPE**

Some future works which can be included to improve the

model and also to make it more effective in practical cases.

Our future works include applying PCA (Principal Component

Analysis) in active learning process to fully automate data

labeling process with less assistance from the oracle. The

model can be incorporate with programs that can interact with

customer seeking a score of a particular product. As we used a

large scale dataset we can apply the model on local market

sites to get better accuracy and usability. And lastly we will

try to continue this research until we generalize this model to

all kinds of text based reviews and comments.

**11 .BIBILIOGRAPHY**

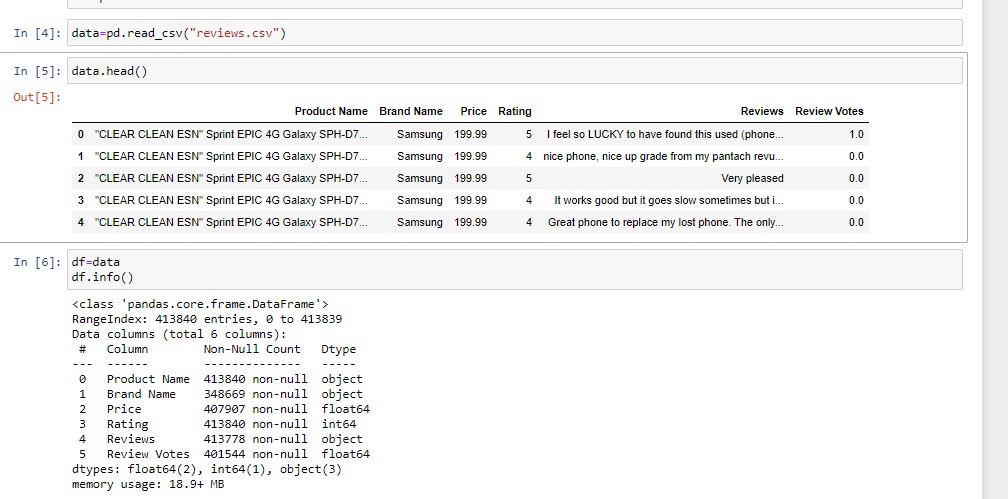
<https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews>

<https://www.kaggle.com/yaowenling/amazon-cell-phone-review-nlp>

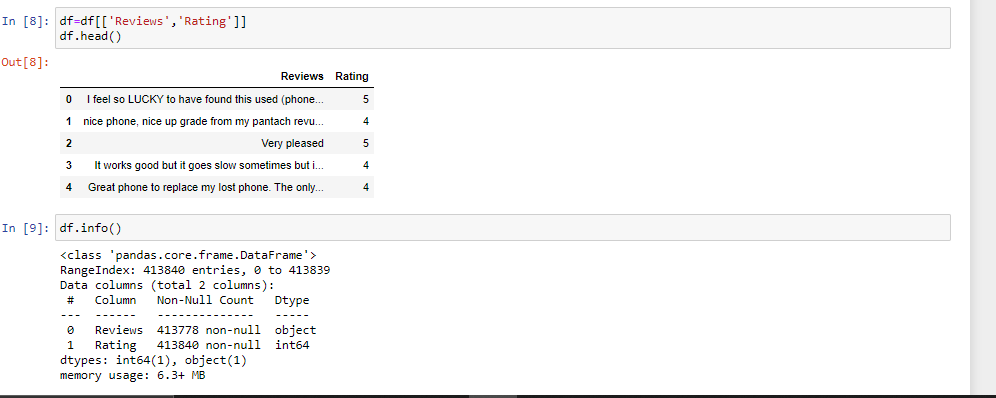
<https://github.com/grikomsn/amazon-cell-phones-reviews>

# 12. APPENDIX

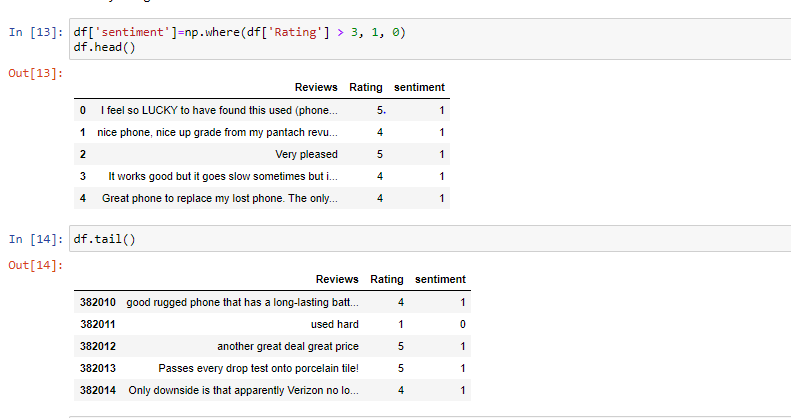
**Source Code:**

Imorting Dataset:

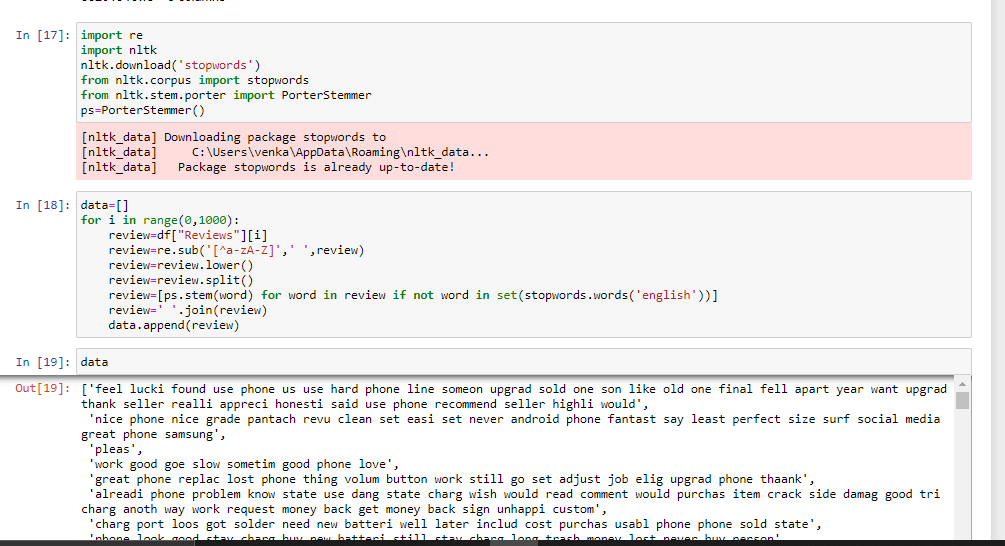
Selecting ecessary rows:

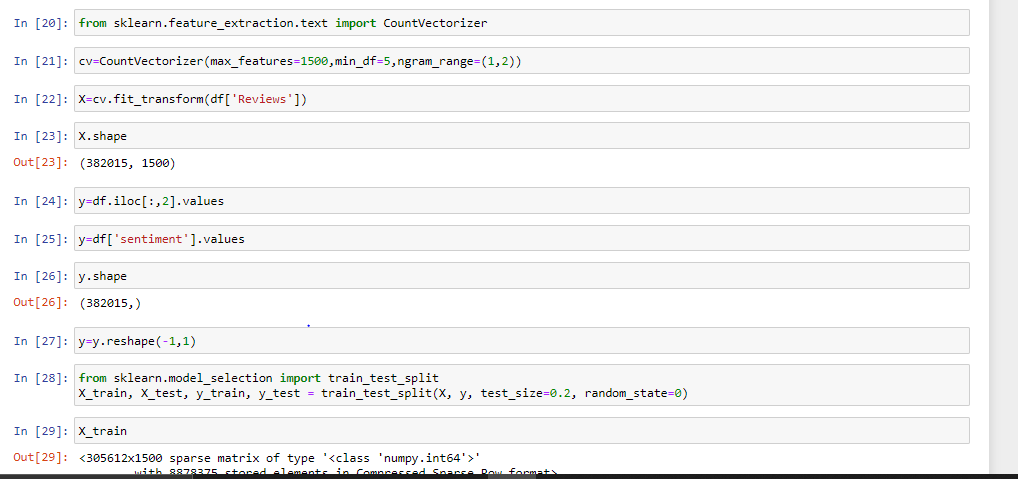


Creating dependent variable:

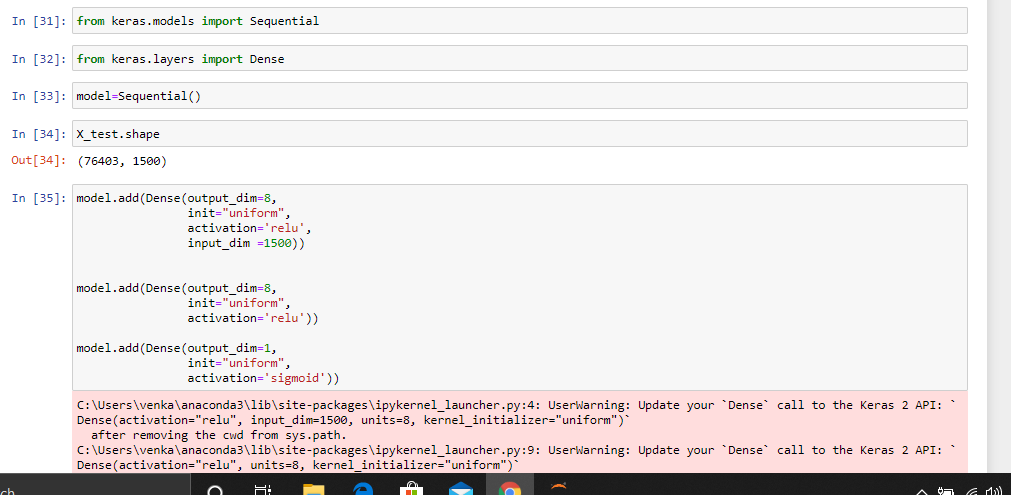


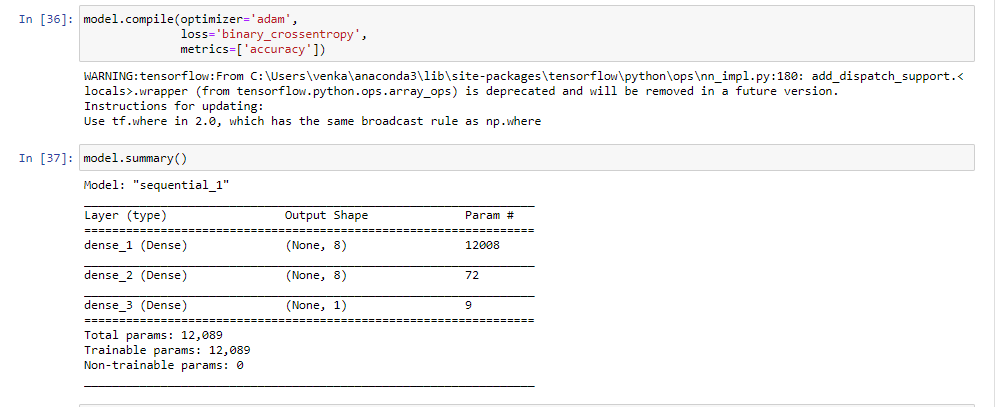
Importing packages and Text preprocessing:

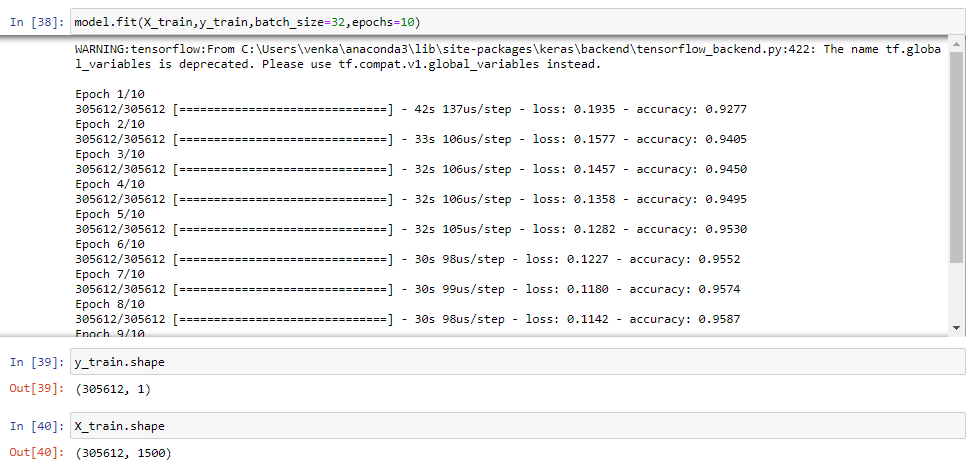




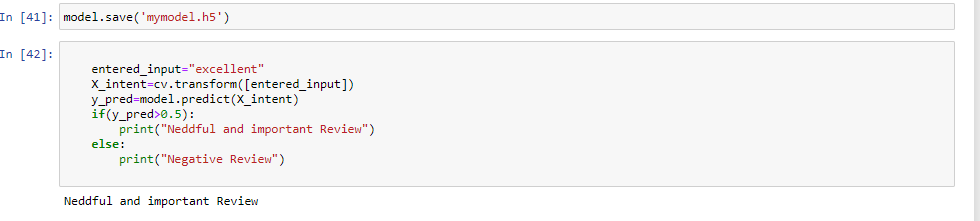
Importing modules and Model building:



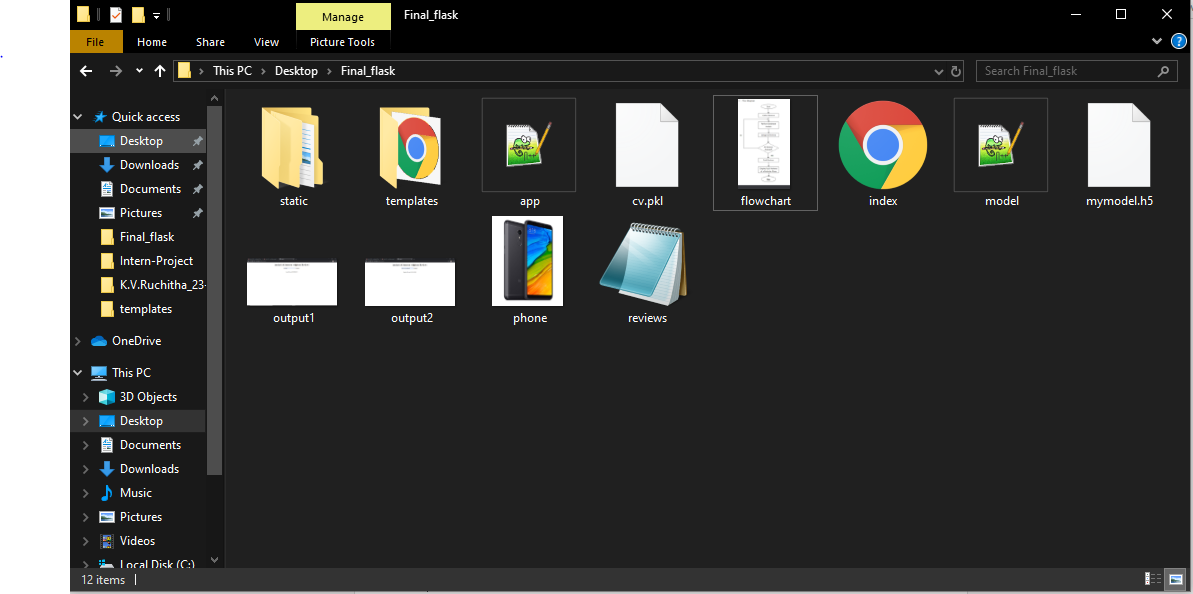


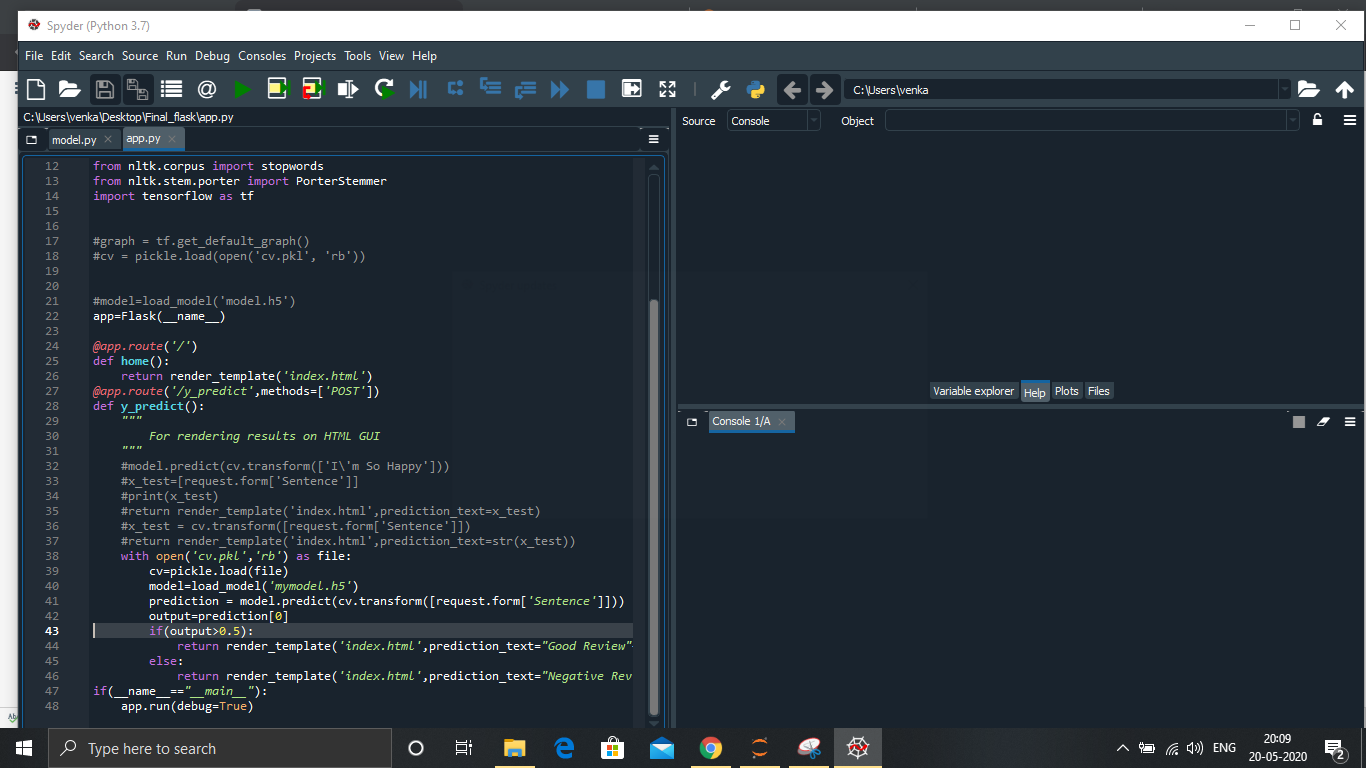


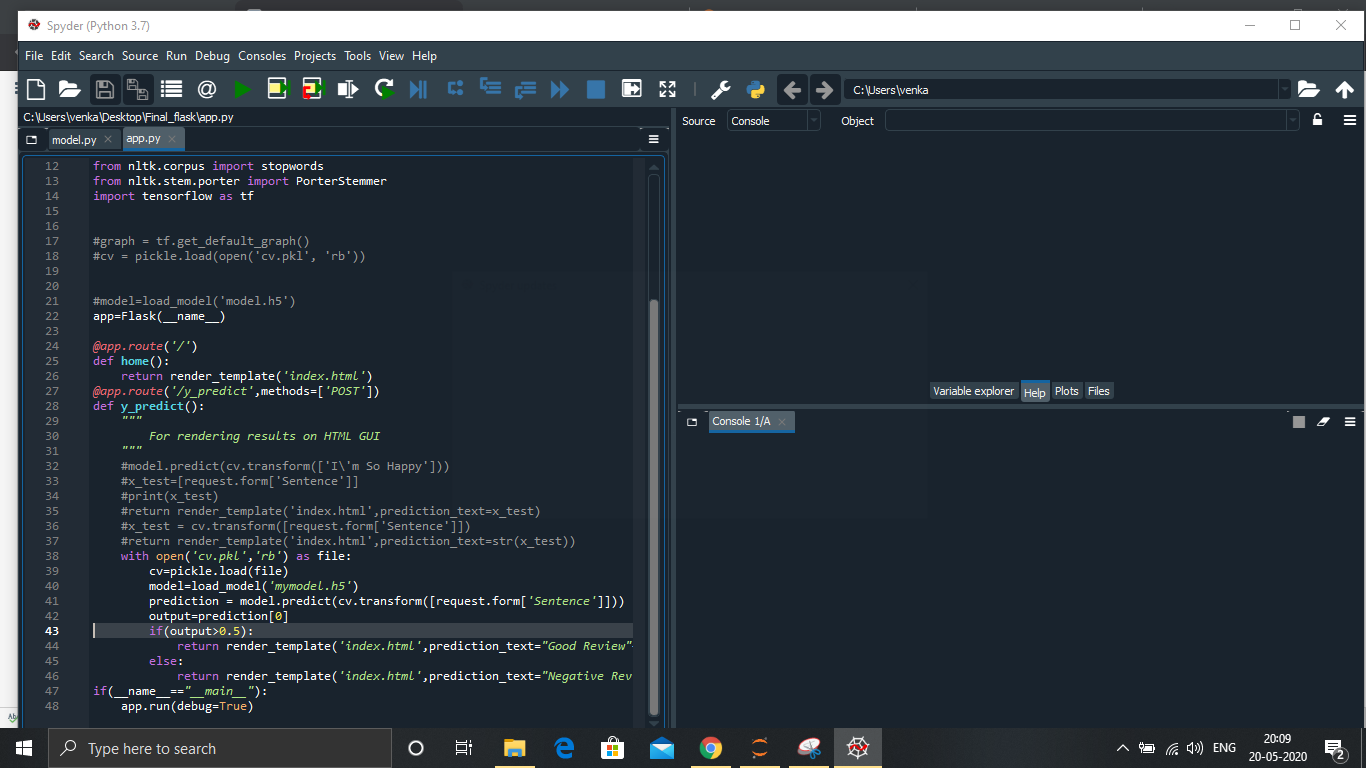
Model saving and Prediction:



Flask Integration:







Ouputs:

