**Exploring the Effectiveness of LSTM Model in Sentiment Analysis**

**Haidy Haitham 212539**

**Mahmoud Hossam 222125**

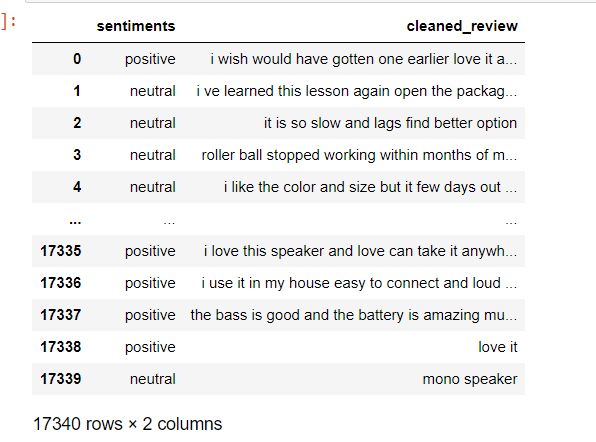
# INTRODUCTION

This study demonstrates the effectiveness of deep learning in handling real-world problems such as sentiment analysis by utilizing NLP techniques and the LSTM model. The novel method to text data preparation has improved the model's accuracy and performance, making it a valuable tool for businesses and organizations wanting to monitor consumer feedback and sentiment.This report describes the use of a Long Short-Term Memory (LSTM) neural network in the implementation of a sentiment analysis project. The project includes data preparation , model architecture development, model performance evaluation, and model deployment. The project's goal is to forecast whether customer evaluations will be positive, negative, or neutral.Word2vec was used as a representation after preprocessing the data.

# Data preparation

# The initial step in the project was data preparation. The data was installed from Kaggle . The pandas package was used to read the dataset from a CSV file. The data was then preprocessed, with links, special characters, and digits removed. The NLTK library was also used to eliminate stop words from the text. The preprocessed data was then divided into features (X) and targets (y), with the target representing the review's emotion (positive, negative, or neutral). After preprocessing the data, Word2vec was employed as a representation.Then dividing the data into training and testing using train\_test split .

The data after dropping the unnecessary Features



# 

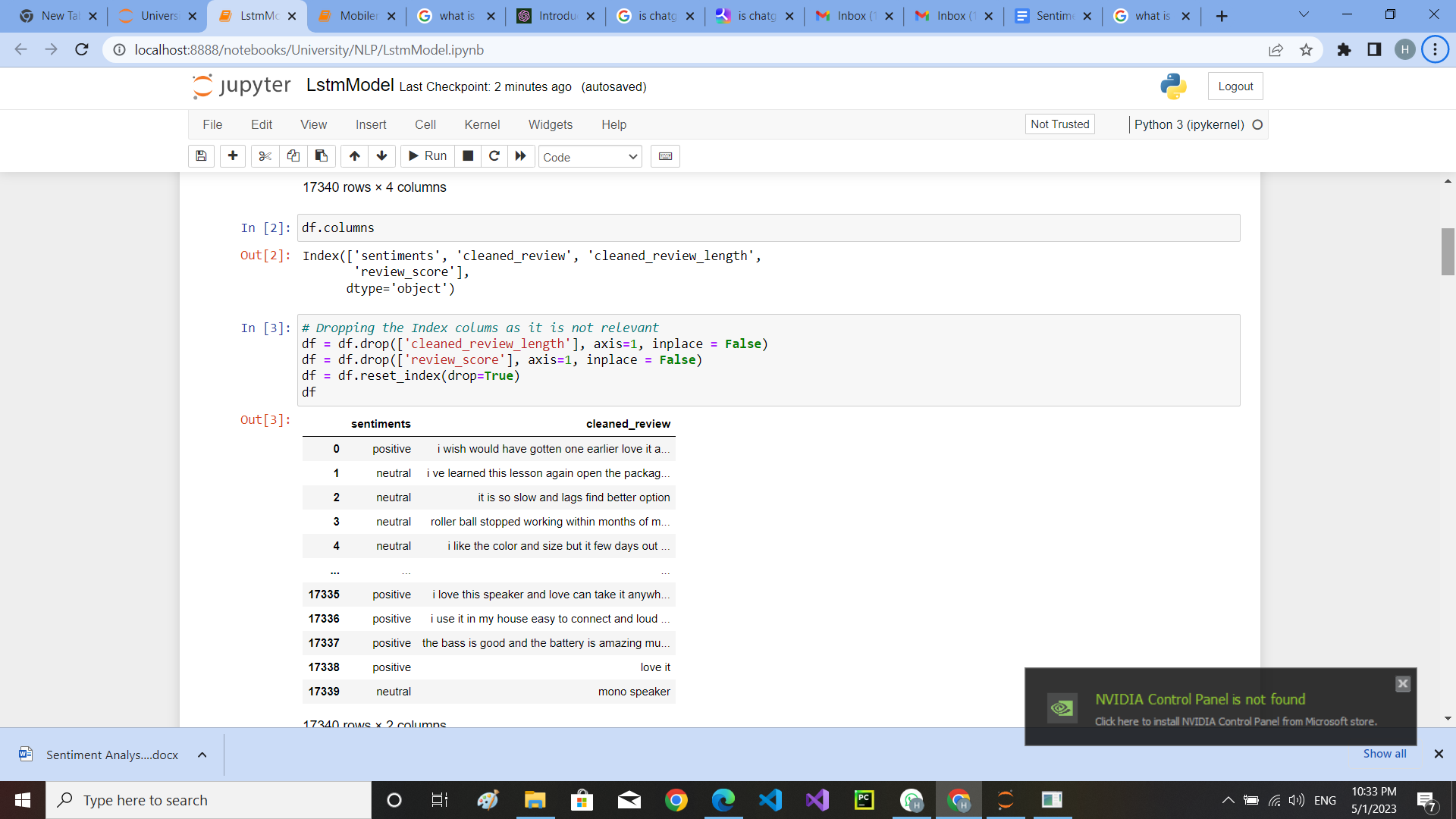
# 

# 

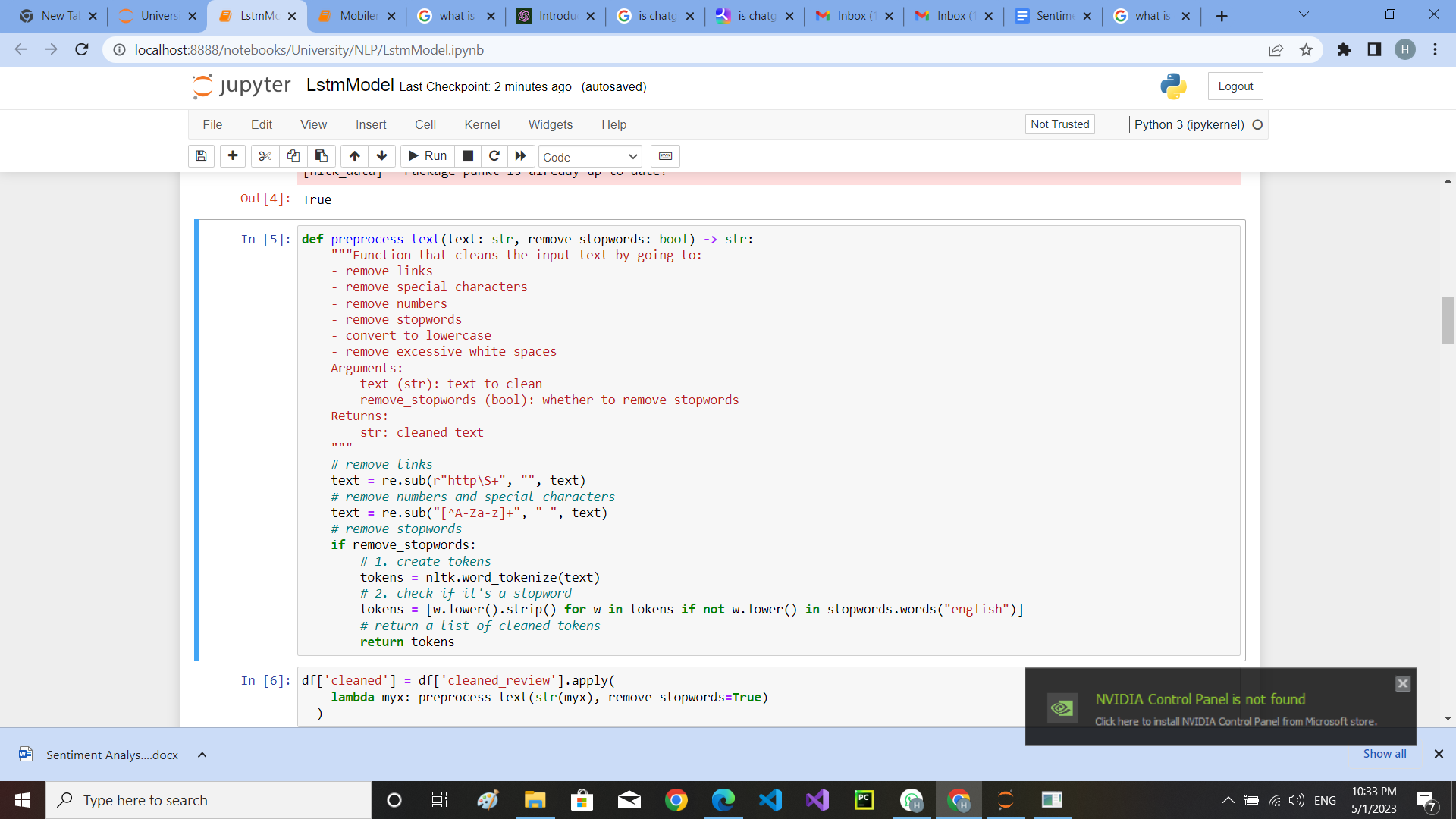
# 

# 

# Code of preparation :

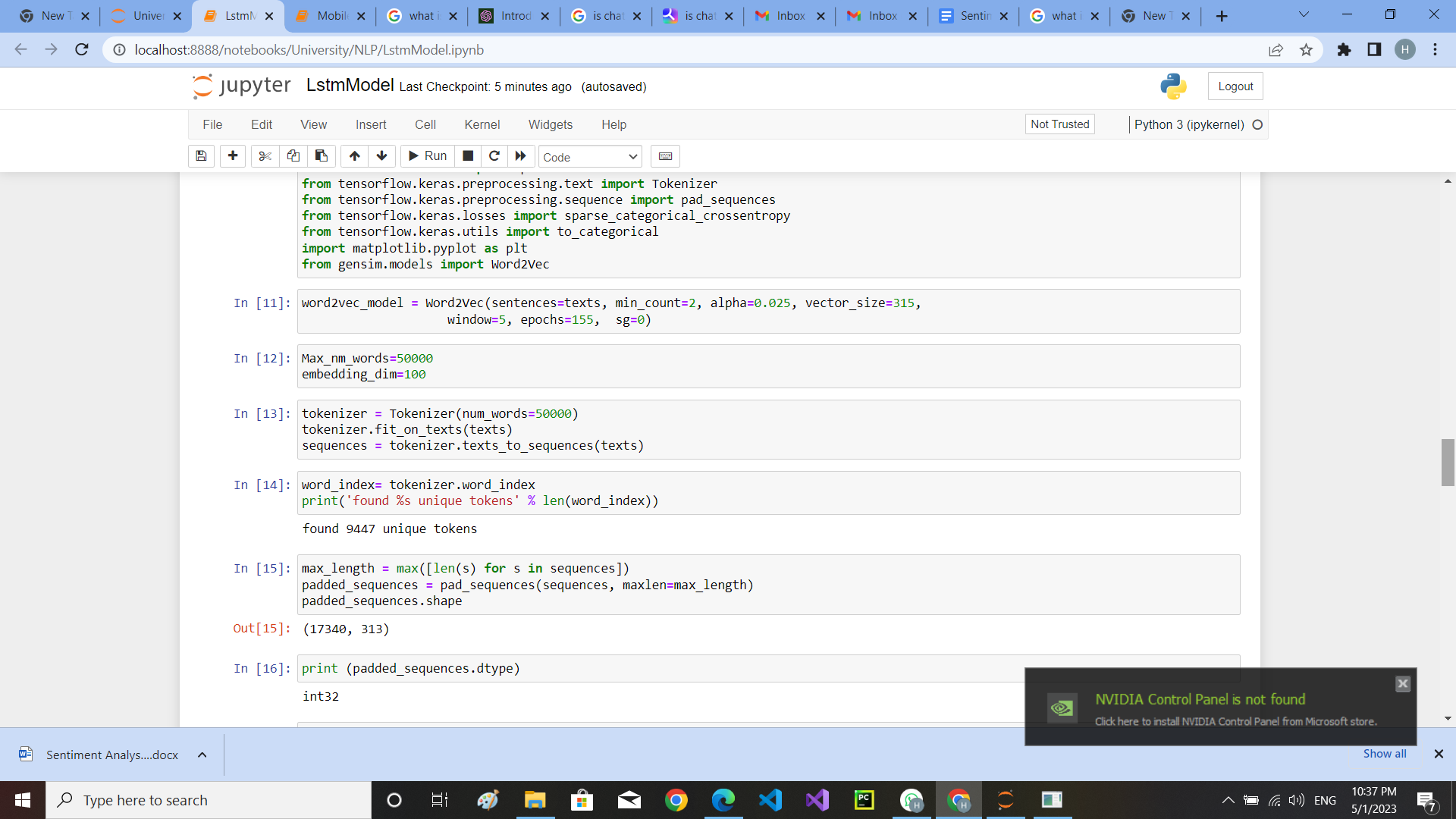


# Pre-processing:

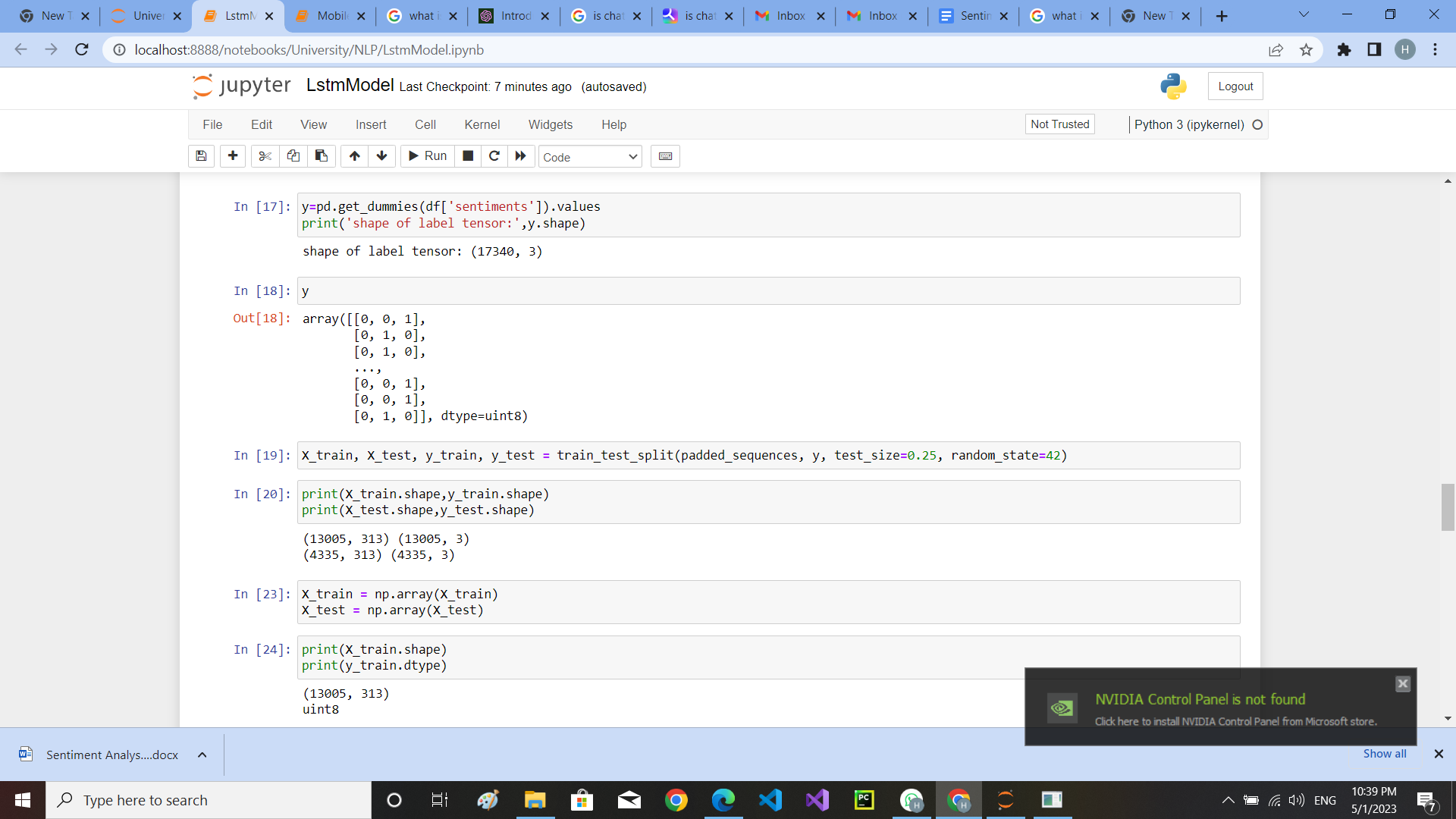


# 

# Word2vec and preparing the variables for the model architecture

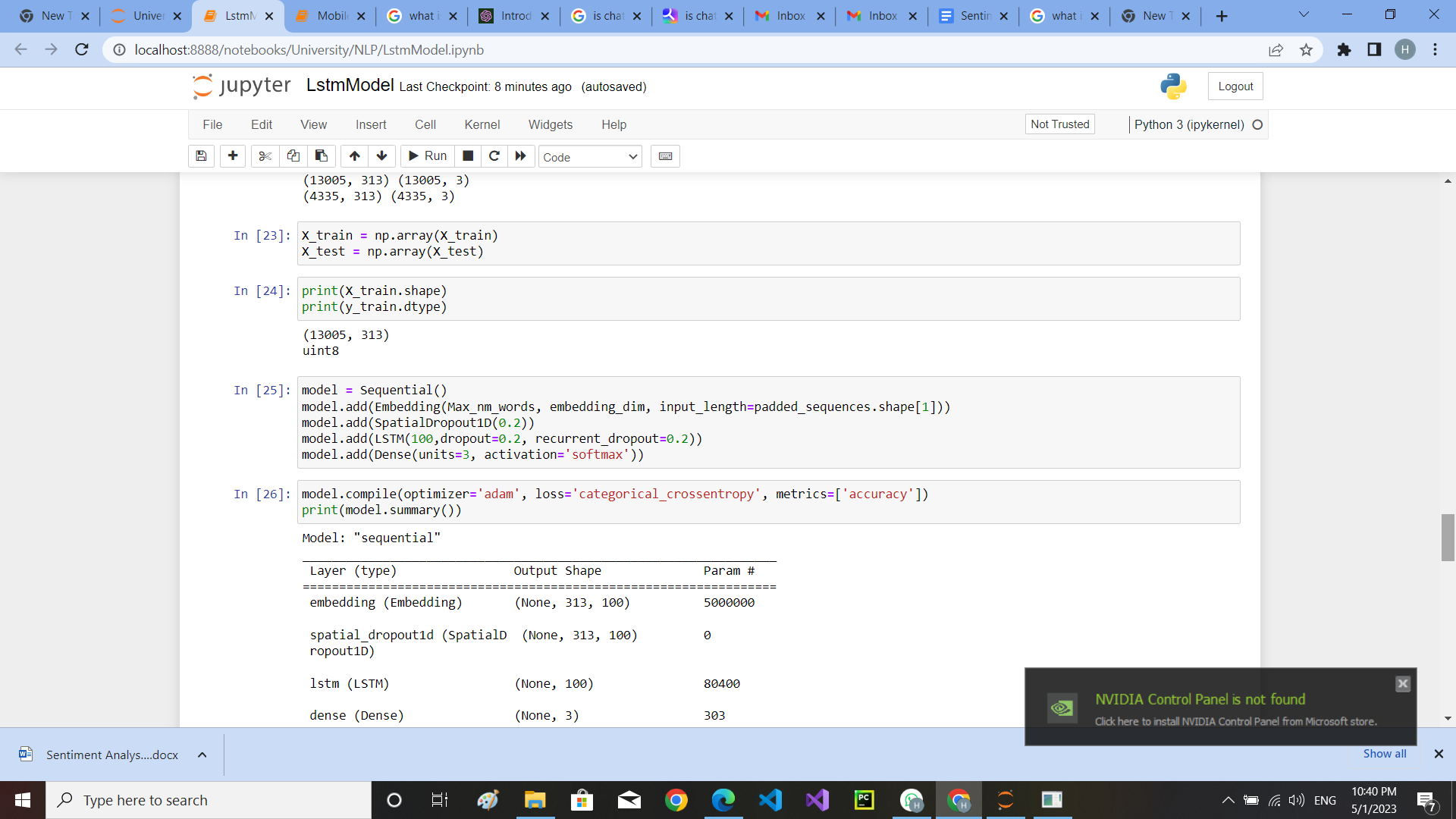


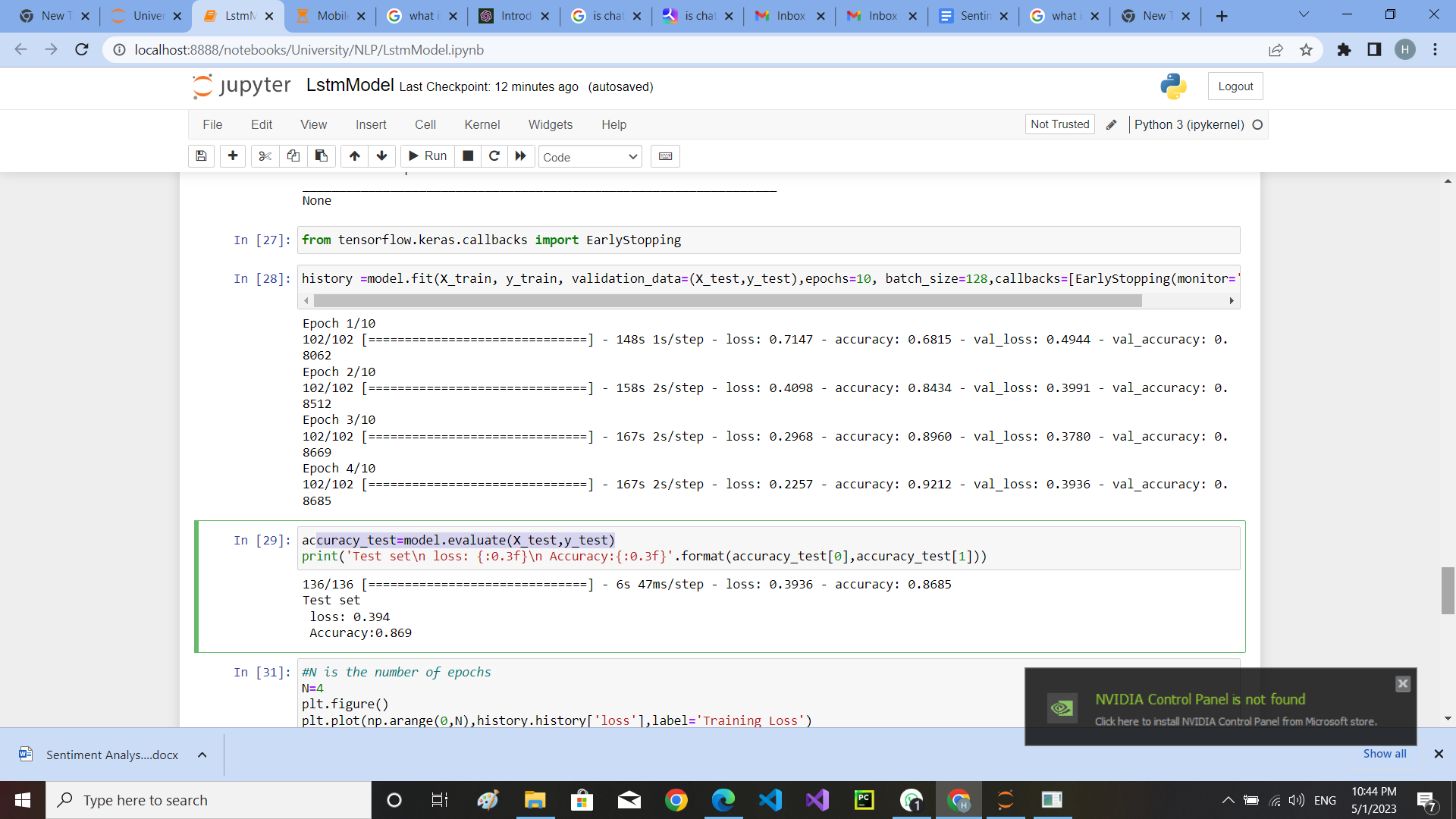
# Splitting the data



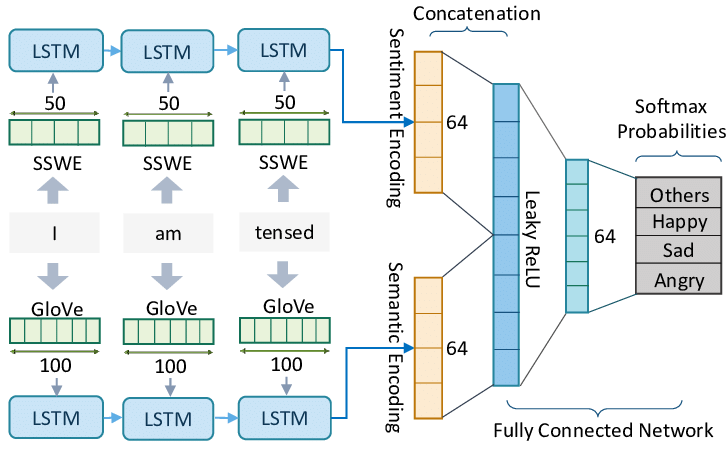
# Model Architecture

The Keras library was used to create the LSTM model. An embedding layer, a spatial dropout layer, an LSTM layer, and a dense layer with a softmax activation function comprised the model. The embedding layer mapped the text data to a high-dimensional vector space, while the LSTM layer examined the text data's sequential nature. The dense layer classified the sentiment of the review into one of three categories (positive, negative, or neutral), while the spatial dropout layer prevented overfitting.





The architecture of Sentiment and semantic LSTM(SSLSTM)MODEL



# Model Evaluation

The model was trained with training data and tested with test data. To avoid overfitting, the model was trained for 10 epochs before being stopped. Due to an early stop, the end epoch was at epoch 4/10. The model obtained 92.12% accuracy on the training data and 86.85% accuracy on the test data. The loss and accuracy for the training data were 0.2257 and 0.9212, respectively, while 0.3936 and 0.8685 for the test data. Predictions were produced on the test data to obtain the evaluation metrics, and classification\_report from sklearn.metrics was used to obtain the precision, recall, f1-score, and support for each class. The model has great precision, recall, and f1-score for all classes, according to the results.

| Class | Precision | Recall | F1-Score | Support |
| --- | --- | --- | --- | --- |
| Positive | 0.71 | 0.72 | 0.72 | 398 |
| Neutral | 0.85 | 0.81 | 0.83 | 1579 |
| Negative | 0.91 | 0.93 | 0.92 | 2358 |
| Overall (Accuracy) | 0.87 | 0.87 | 0.87 | 4335 |

# 

# 

# 

# 

# 

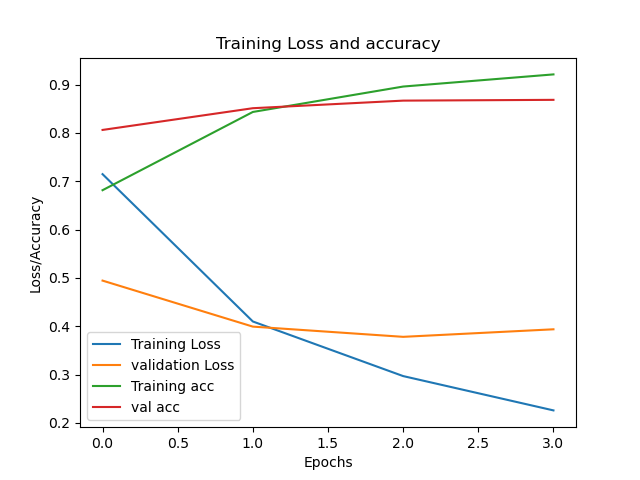
# 

# 

# 

# 

# The diagram that shows the Training loss and accuracy and the validation loss and accuracy



# Deployment of the project

Model Selection and Training: During this stage, the Model was selected and trained on the preprocessed dataset. Neural Networks such as LSTM was chosen for sentiment analysis.

Tuning Hyperparameters: After selecting a model, it was necessary to tune its hyperparameters to improve its performance. This is accomplished through the use of techniques such as Random Search.

Model Evaluation: After training and optimizing your model, It was a must to test it on a holdout dataset. Accuracy, precision, recall, F1-score, are common sentiment analysis evaluation measures.

# 

# CONCLUSION

# In conclusion, the LSTM model was able to accurately classify the sentiment of customer reviews into one of the three categories (positive, negative, or neutral). The model achieved an accuracy of 86.85% on the test data, which was a good result.