INTENT ANALYSIS

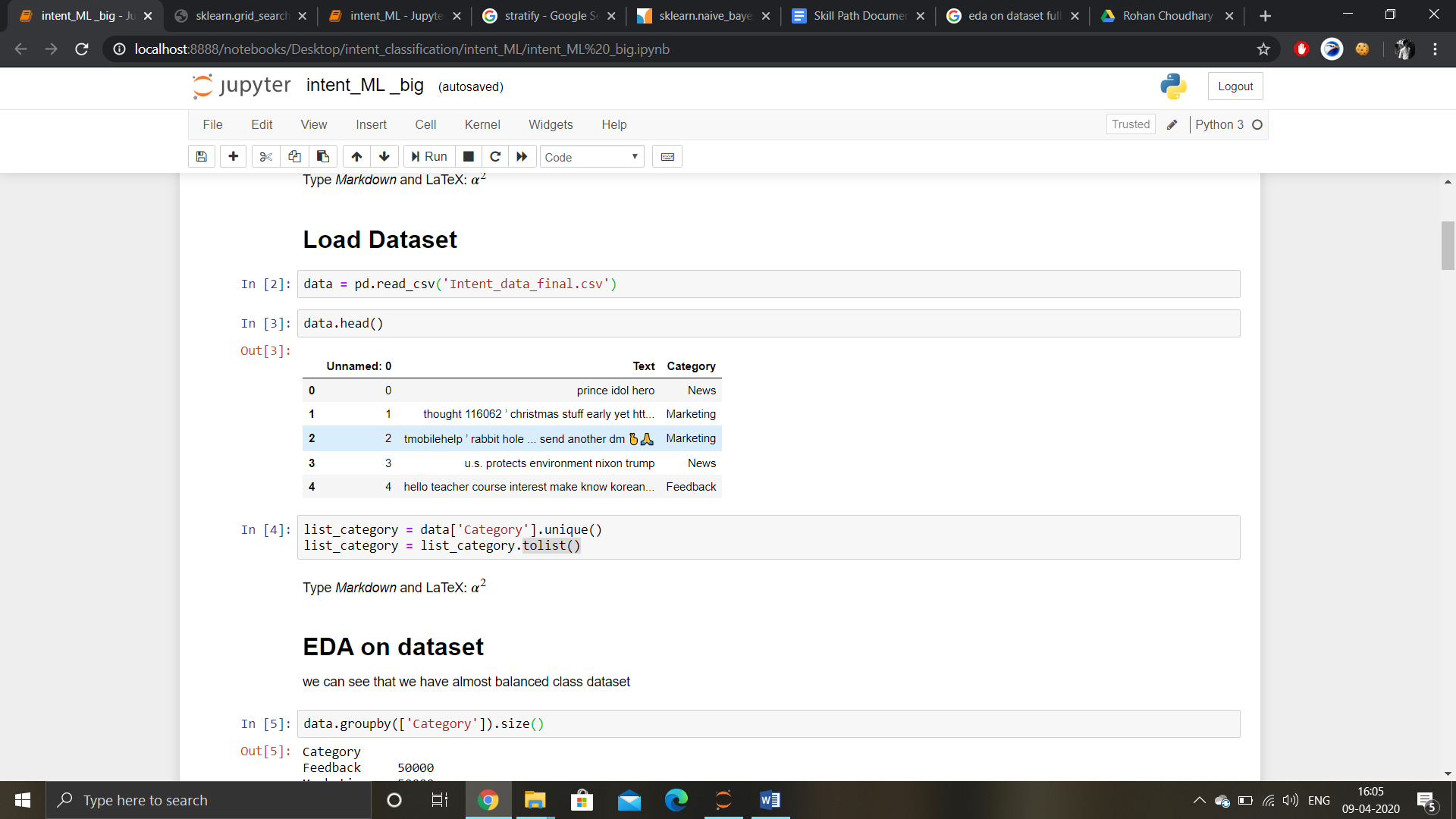
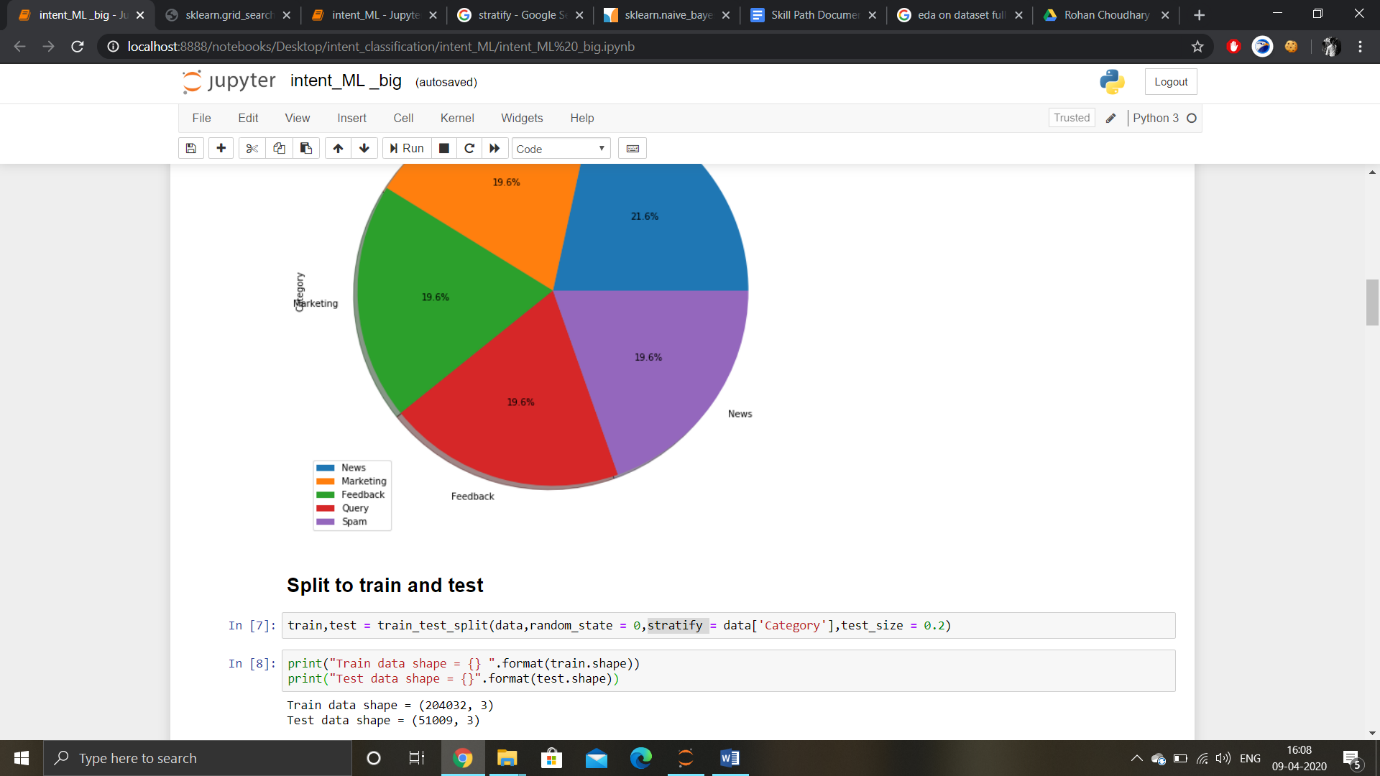
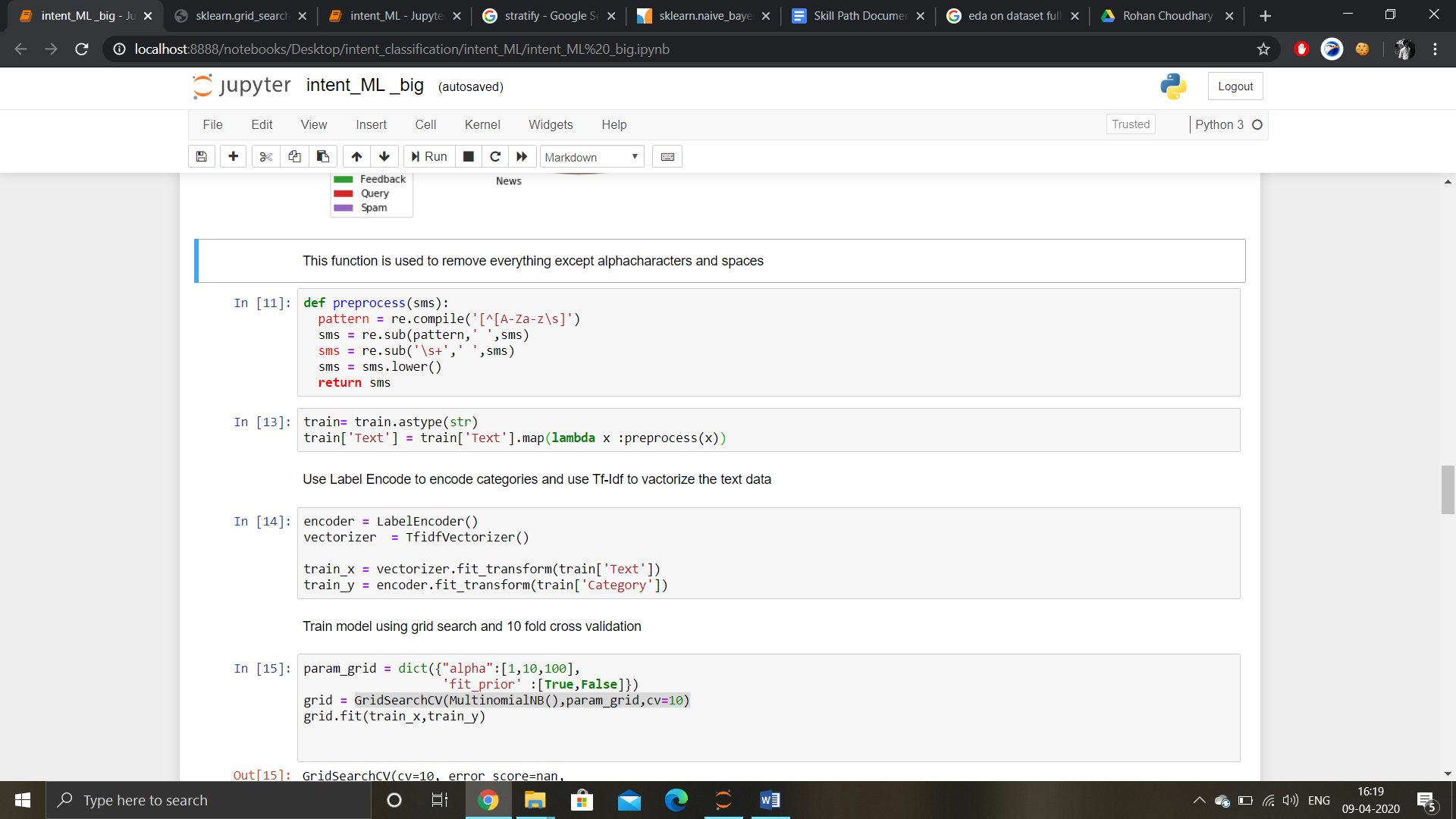
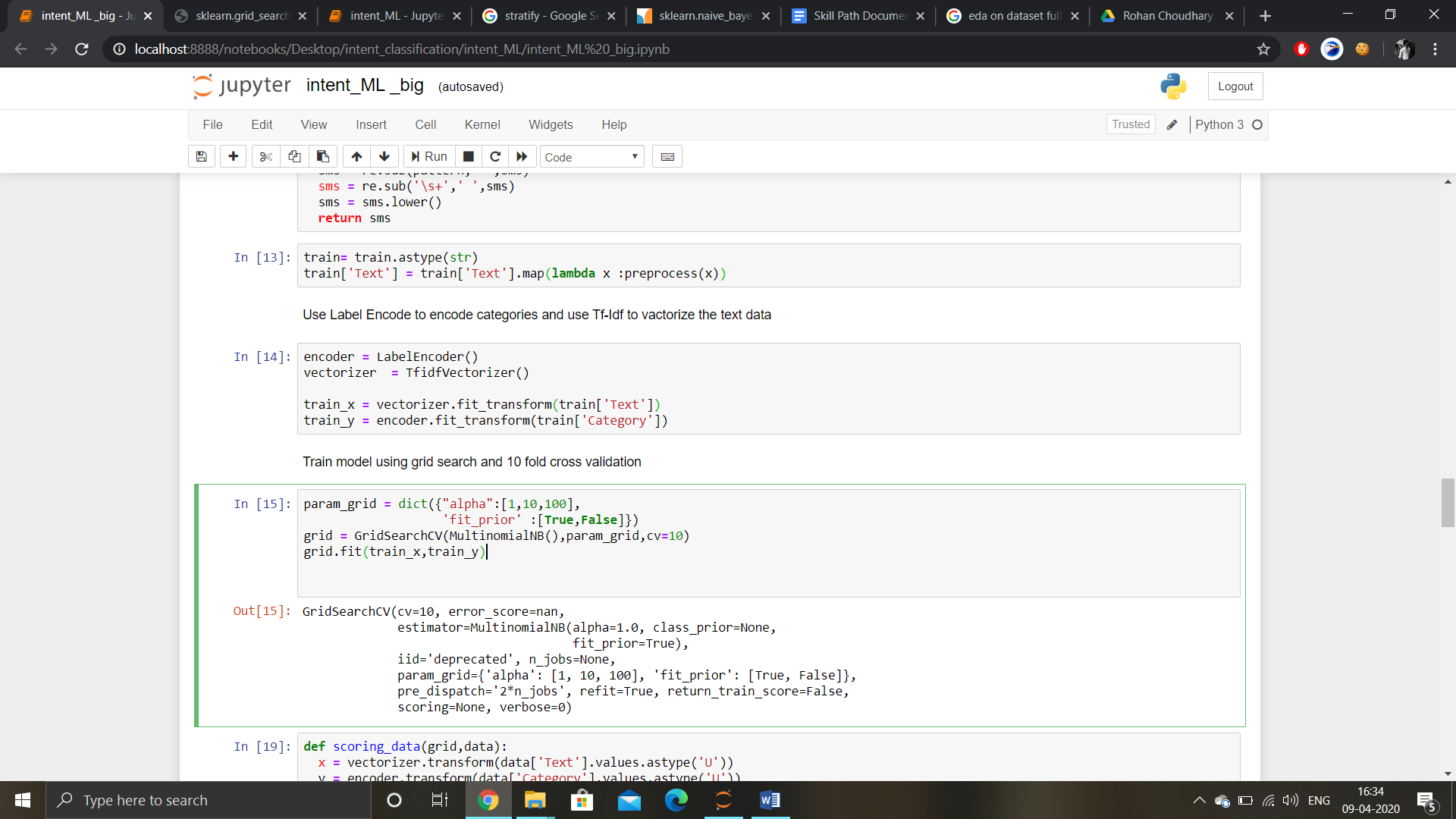
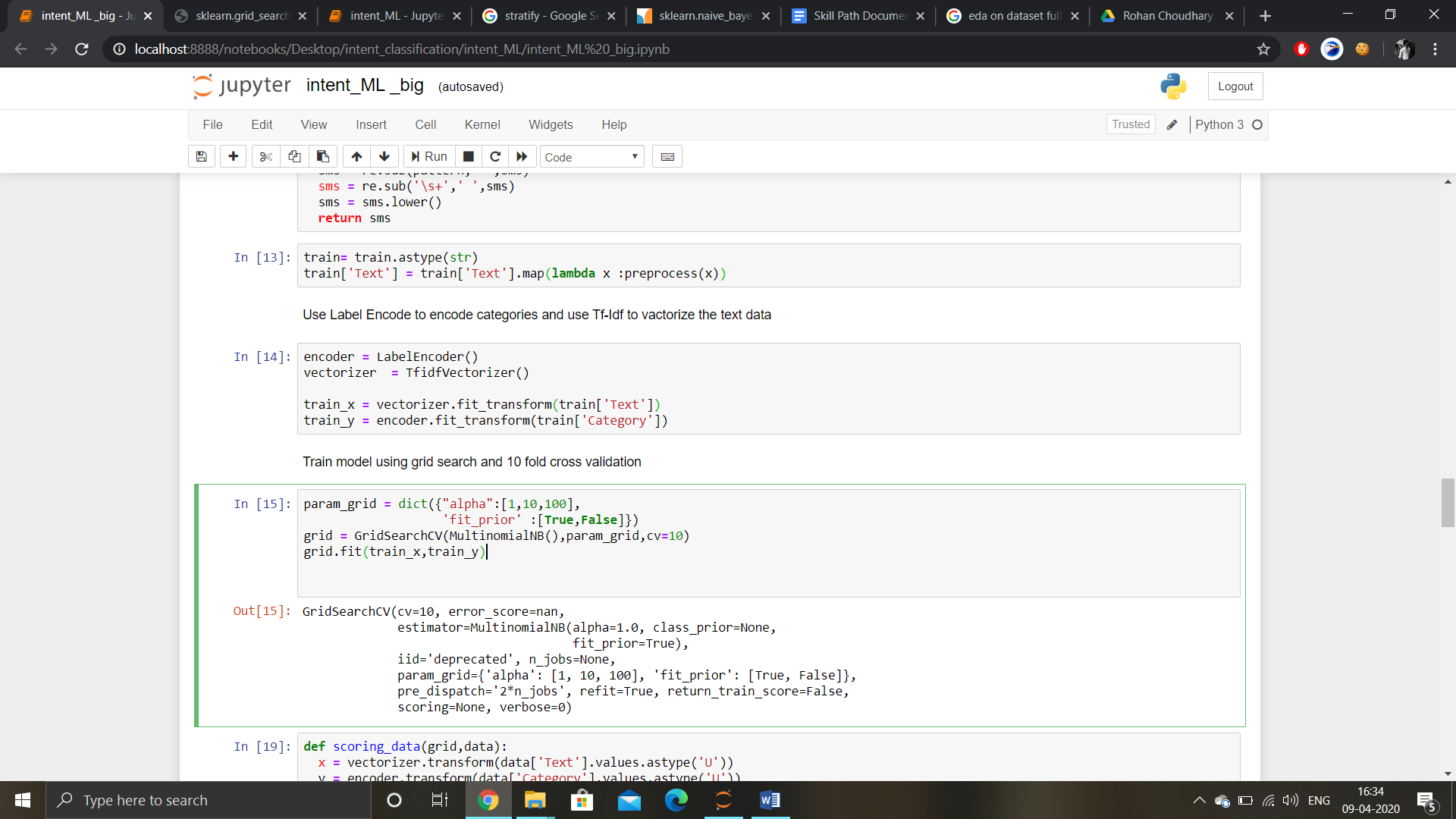
Authored by: Rohan Choudhary

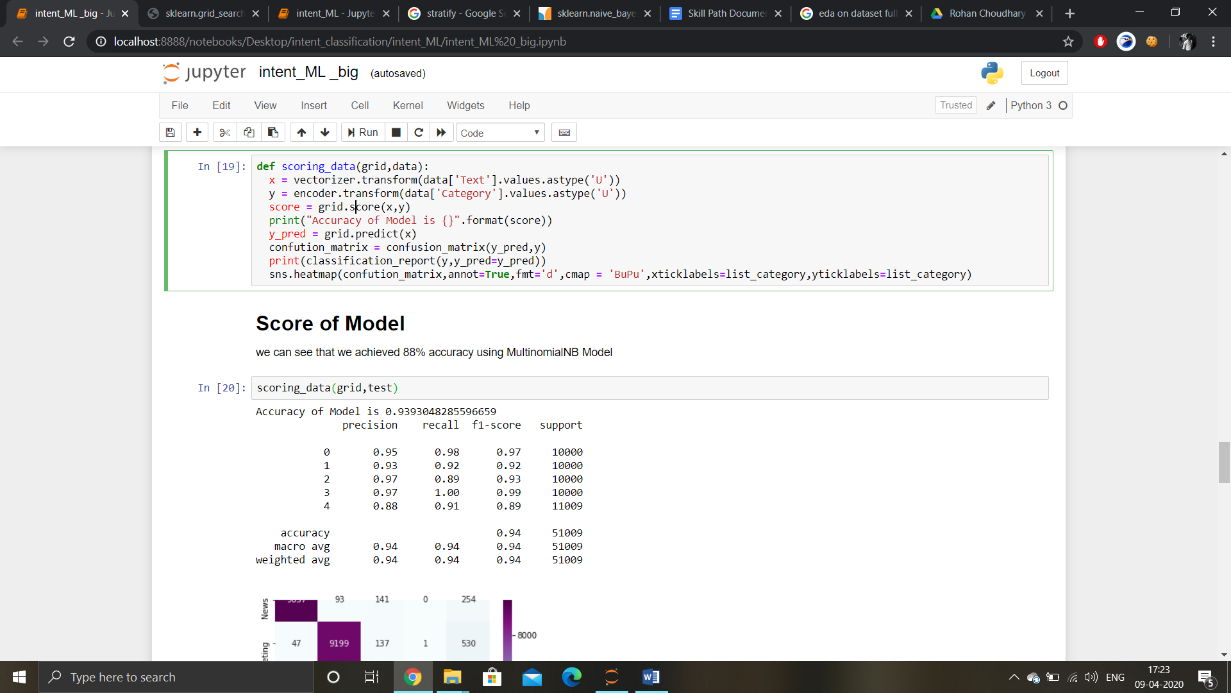
Objective: Try to predict the intention behind a comment broadly into categories like feedback, query, marketing, spam and news.

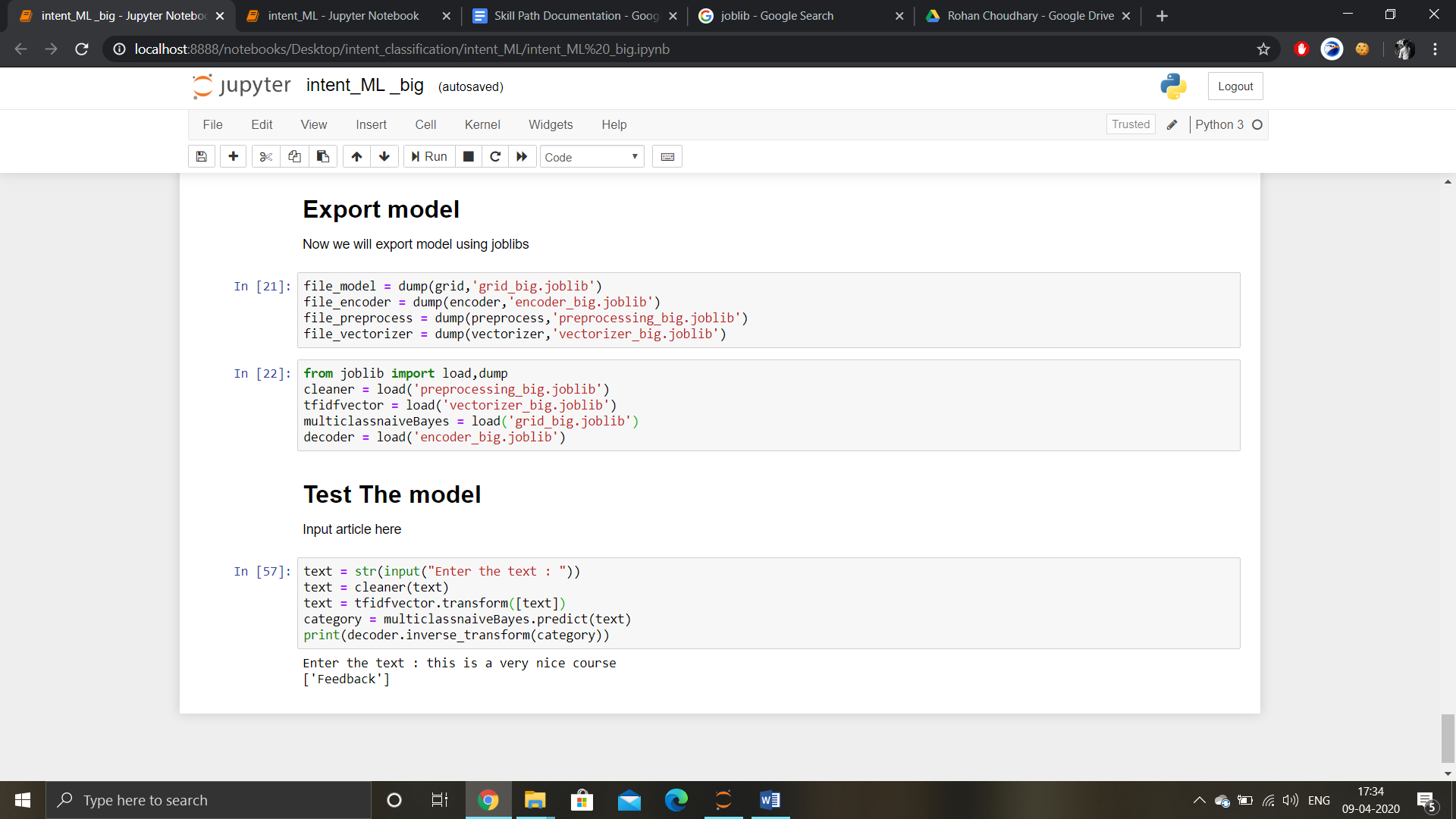
Skill Path: Python, Machine Learning, Natural Language Processing.

Note: Various ML models were used to predict intents, the one with the highest accuracy has been used in the end and mentioned in this documentation.

Filename: intent\_ML\_big.ipynb

1. The dataset Intent\_data\_final.csv has been used in this script. The dataset contains almost 5000 entries for five intents, namely feedback, query, marketing, news and spam.
2. Exploratory Data Analysis on dataset is done and the dataset is divided into test and training dataset (80:20 ratio).
3. Pre-process function is created to remove everything except alpha characters and spaces from the dataset.
4. Use Label Encode to encode categories and use Tf-Idf to vactorize the text data.
5. Train model using grid search and 10 fold cross validation and Multinomial Naïve Bayes. [GridsearchCV examples](https://www.programcreek.com/python/example/104786/sklearn.grid_search.GridSearchCV). [<Link2>](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)
6. Function: scoring data for studying the accuracy parameters. Achieved 94% accuracy.



1. Now we will export the model using joblibs (**Joblib** is a set of tools to provide lightweight pipelining in Python) and try new examples to test the model prediction.