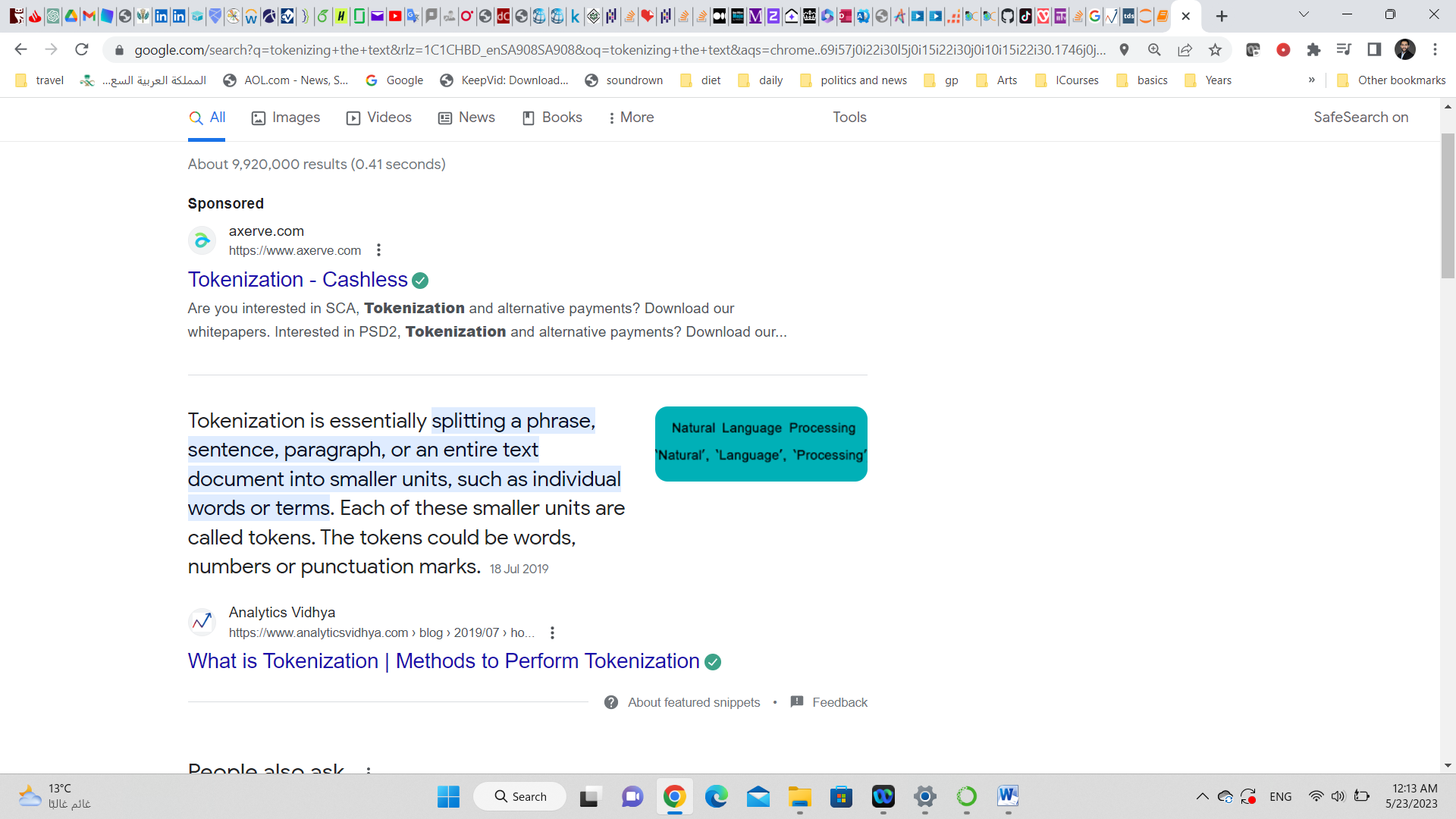
Learn from the task

Web scraping

<https://realpython.com/beautiful-soup-web-scraper-python/>

**Topic Modeling in Python: Latent Dirichlet Allocation (LDA)**

<https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>



# Sentiment Analysis Using Python

<https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/>

# Sentiment Analysis using TextBlob

<https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>

<https://stackoverflow.com/questions/46764674/sentiment-analysis-on-dataframe>

Important for data cleaning for text:

# Text Cleaning for NLP: A Tutorial

<https://monkeylearn.com/blog/text-cleaning/>

**models**

<https://scikit-learn.org/stable/supervised_learning.html>

random forest

<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

Random Forest Classification with Scikit-Learn

<https://www.datacamp.com/tutorial/random-forests-classifier-python>

# Understand Random Forest Algorithms With Examples (Updated 2023)

<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=One%20of%20the%20most%20important,for%20classification%20and%20regression%20tasks>.

# [How can I fit categorical data types for random forest classification?](https://datascience.stackexchange.com/questions/26283/how-can-i-fit-categorical-data-types-for-random-forest-classification)

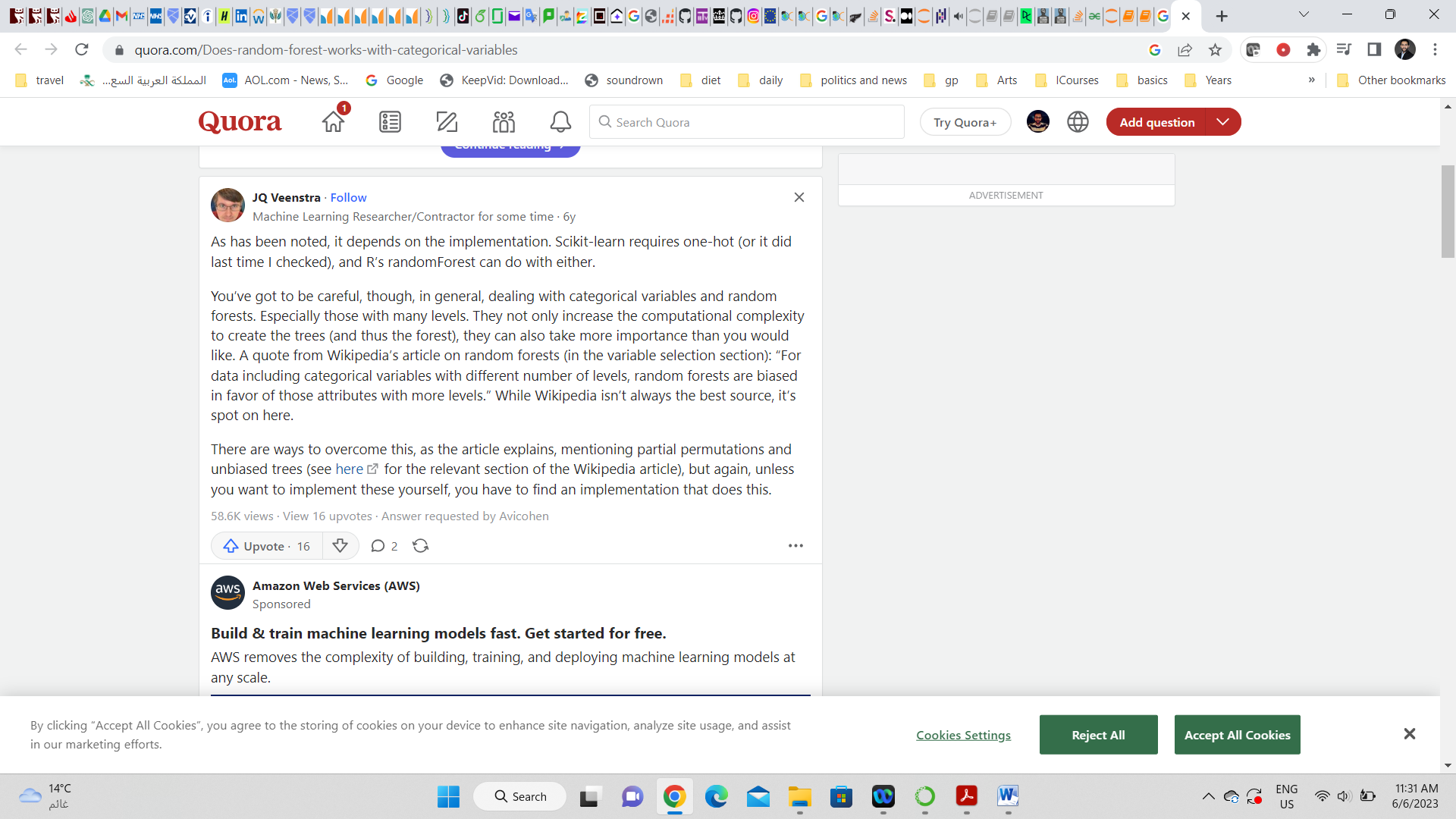
<https://datascience.stackexchange.com/questions/26283/how-can-i-fit-categorical-data-types-for-random-forest-classification>

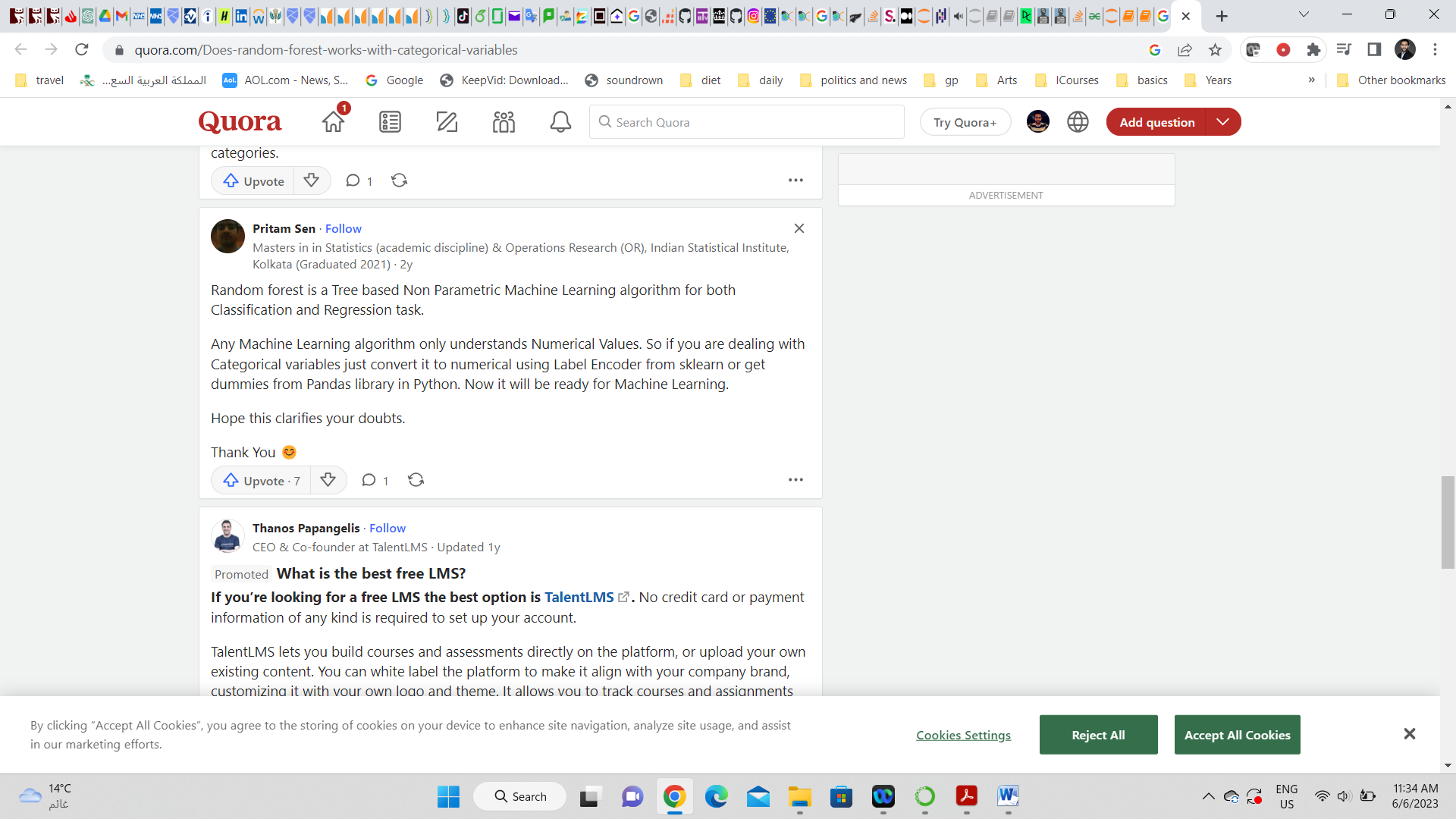
# Distributed Random Forest (DRF)

<https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/drf.html>

good notes:

<https://www.quora.com/Does-random-forest-works-with-categorical-variables>

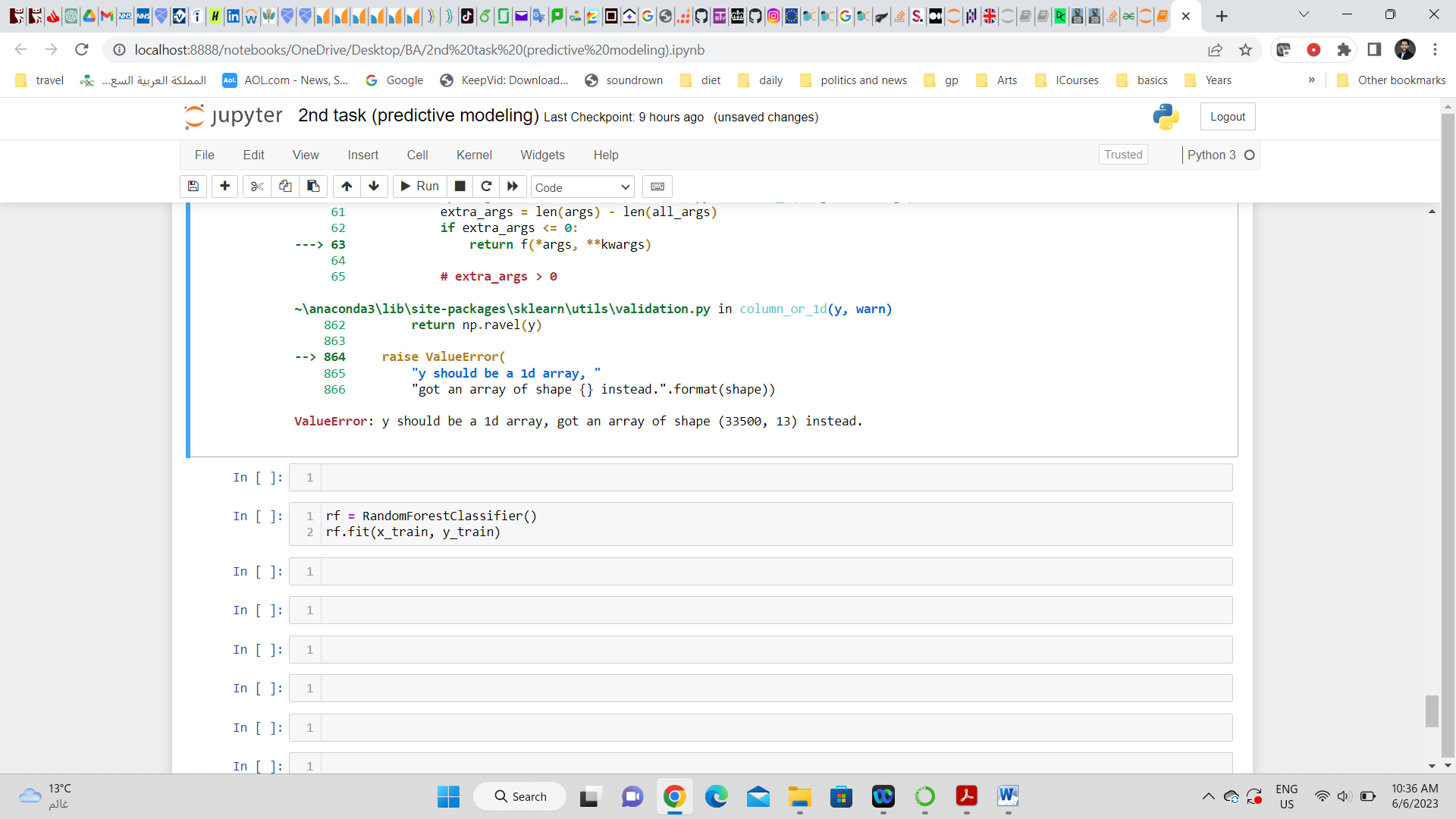




# 10 - Categorical Variables and Random Forest in 9 minutes

<https://www.youtube.com/watch?v=jLXgtr3jrRU>

ERROR:



This error shows that Labelencoder is either for the target only to convert it from categorical to numbers or you need to select only 1d (one column) feature to use LabelEnoder on it

Progress:

I need to convert categorical coulmns to numerical and get rid of high levels columns (that contains numerous categorical values that will make one hot encoding very big)