Finetuning Language Models-Can I patent this

Introduction

The USPTO is the US Patent and Trademark Office. It is the agency that grants patents to inventors and businesses for their inventions.

I created an app that will accept an input patent application and will return its patentability score. The app will be will be used by patent applicants to determine the patentability of their inventions before they file their patent applications, therefore reducing the workload of the patent examiners.

Milestone 1:Docker setup and installation

Docker is a software platform that allows you to build, test, and deploy applications quickly. Docker packages software into standardized units called containers that have everything the software needs to run including libraries, system tools, code, and runtime.

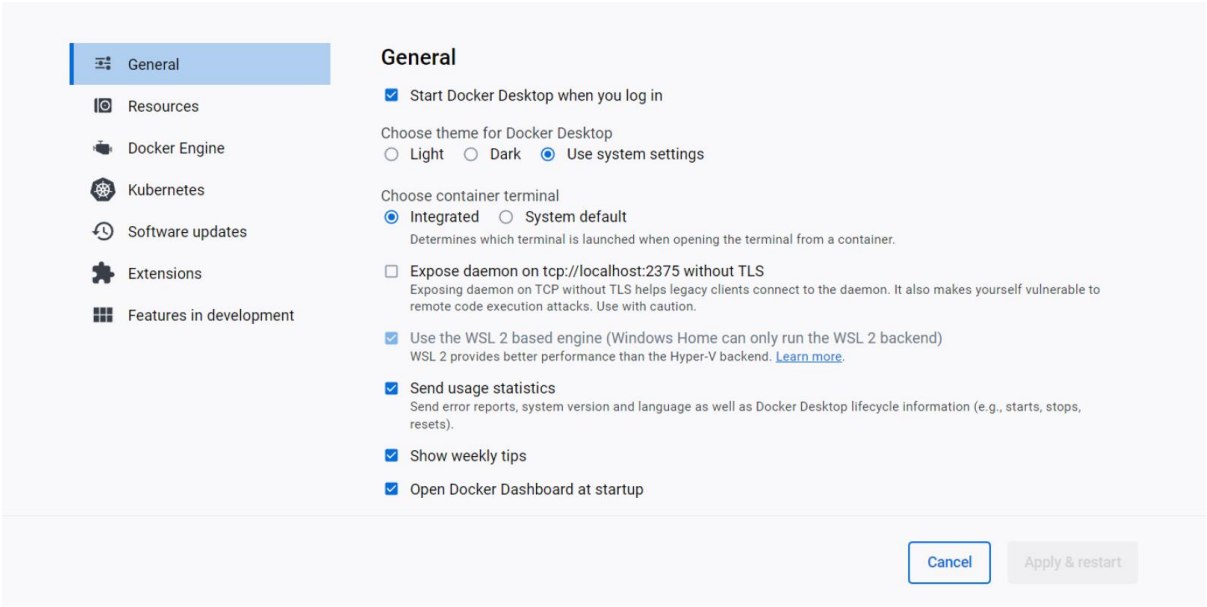
Docker installation instructions :

1. Firstly, Install dockers application(latest) from google.

2. Go to settings, select general and check for the WSL which was selected by default in my system

3. Update the WSL to the latest version.
"https://code.visualstudio.com/docs/remote/wsl-tutorial" and "https://docs.docker.com/desktop/windows/wsl/"

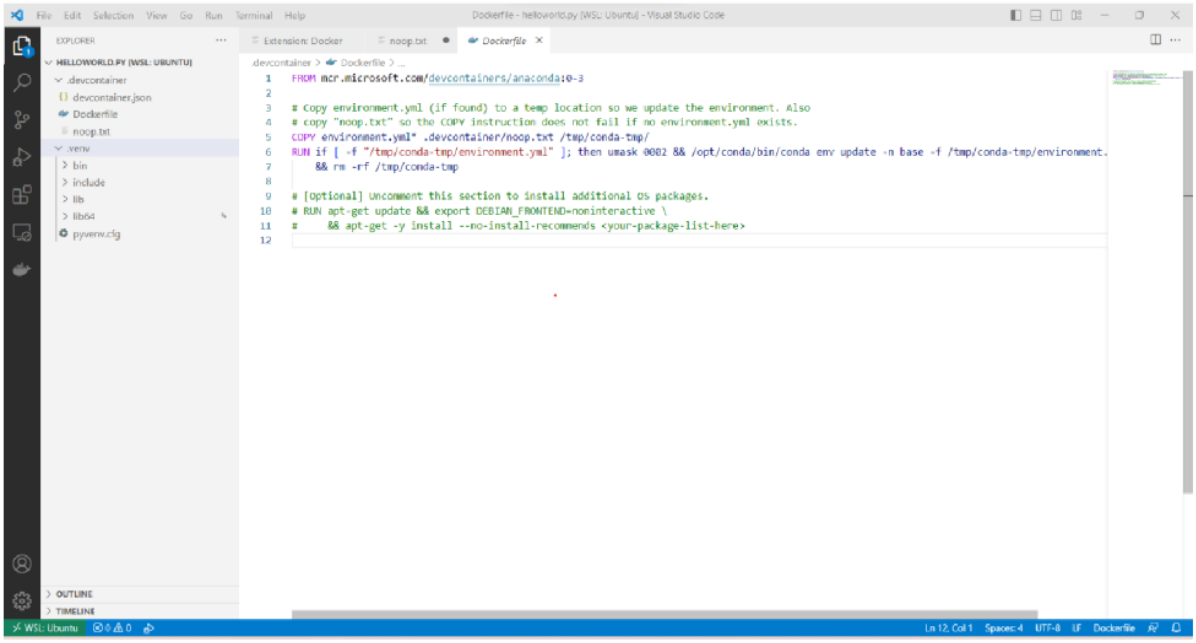
"https://code.visualstudio.com/docs/remote/wsl-tutorial" and "https://docs.docker.com/desktop/windows/wsl/"



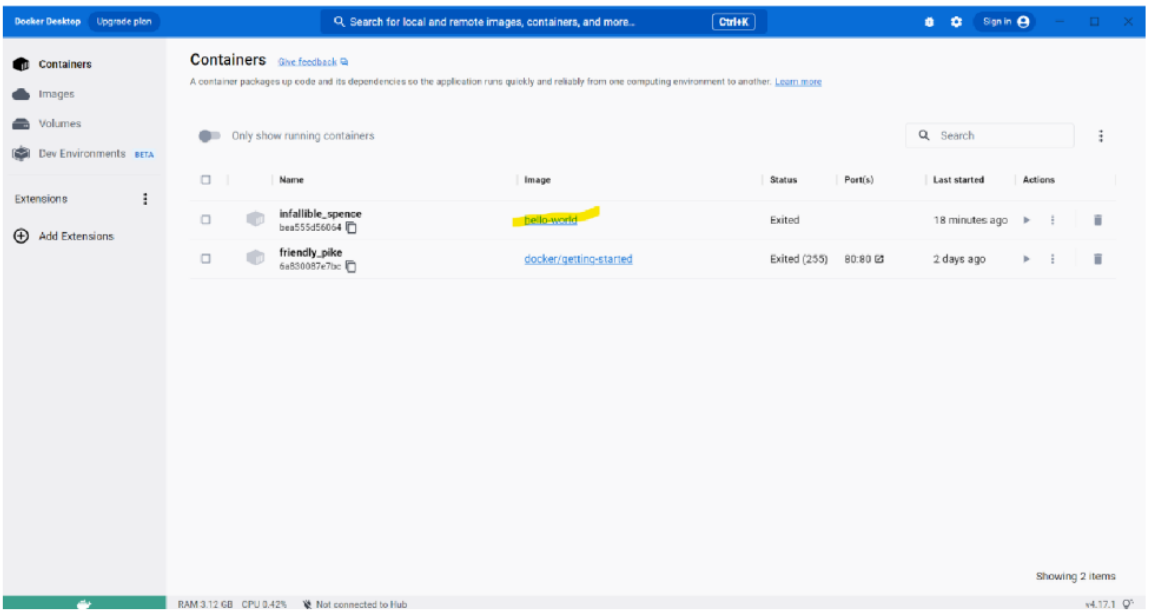
4. Now, install Ubuntu

5. Open the Visual Studio code and download the WSL extension pack and then install remote control development extension pack which sets up a development environment for the system. Click ctrl+shift+p and connect to "WSL : connect to the default distro"

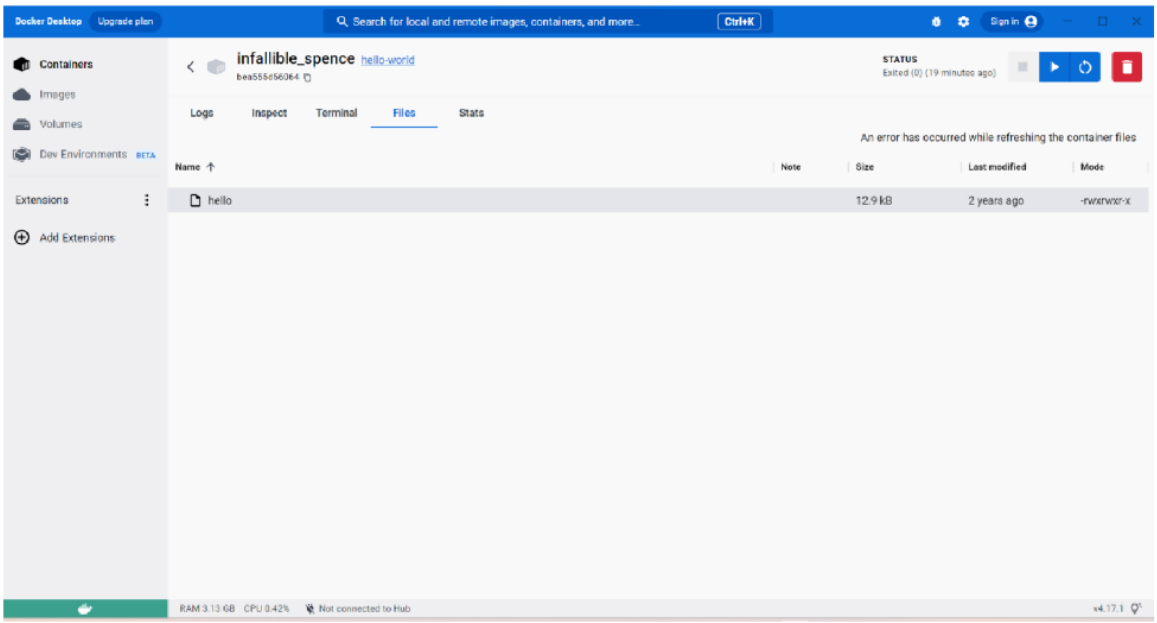
6. Now, open terminal and write "code ."



Below is the container in the docker created with the name hello-world.



Below screenshot depicts the created hello python file in the docker



Milestone 2: Sentiment Analysis App

Here we created a streamlit app and deploy a sentiment analysis app in hugging face model.

Here in the web page use can go select the hugging face model.

The app deployment is in the below link:

https://venkataseetharam-cs-634-spring-2023-projec-streamlit-app-k7p6e8.streamlit.app/

Here you have to give the text input and select a model from the list of models.After that you have to click on the submit.Then it tells whether the sentence is positive or negative.Here I used three models for selection and they are distilbert-uncased,bertbase-uncased and Roberta-base.

Steps followed for creating the app:

First we created requiremets.txt for installing required libraries.

Then the code is written in app.py and pushed into hugging face website.

https://huggingface.co/docs/hub/spaces-github-actions/

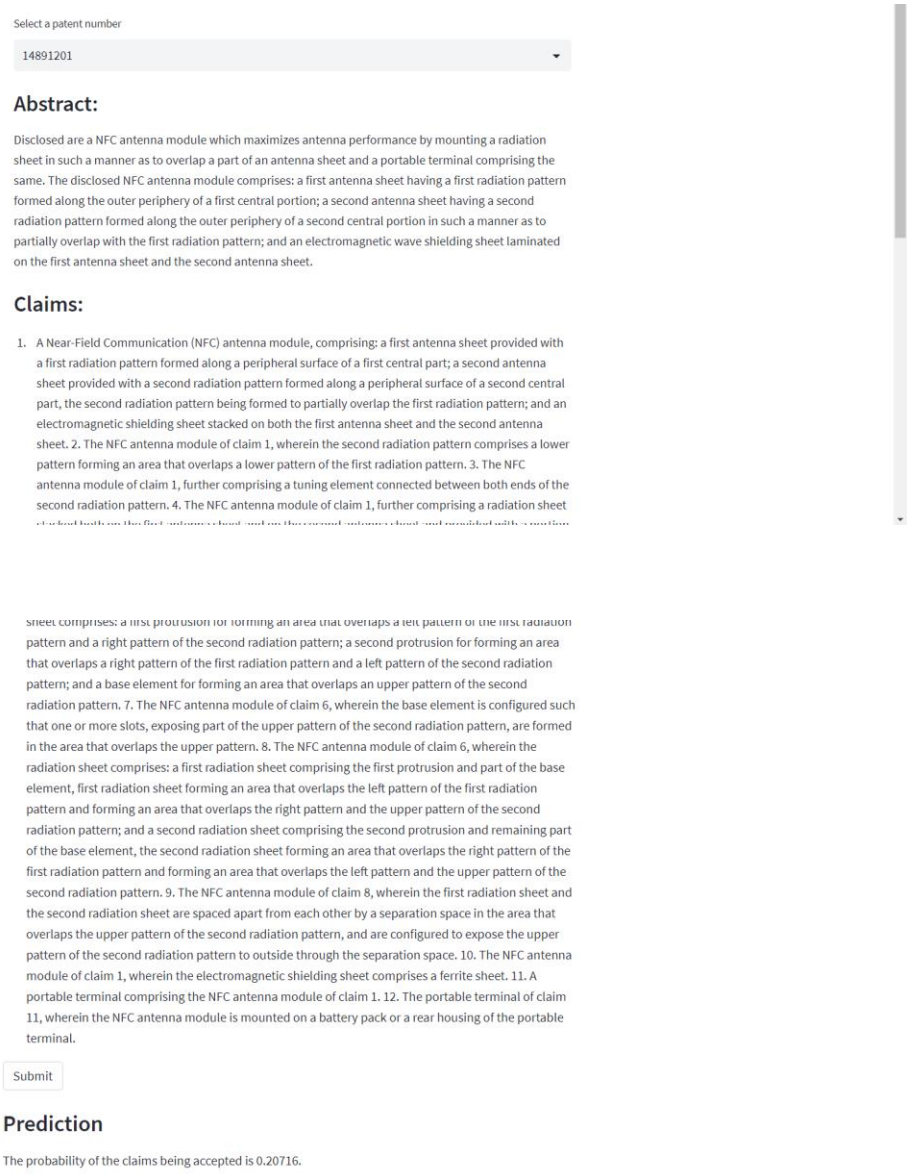
Follow the above link to know more about the models we used

Milestone 3: Finetuning Language Models

Goal: To develop a classifier that analyses the given data and predicts the patentability score to determine how likely is the data going to get the patent right.

Below is the link for the app I designed for patent predictability score:

https://huggingface.co/spaces/venkataseetharam/patentscore



Overview of mile stone-3:

Data:
https://huggingface.co/datasets/HUPD/hupd/blob/main/hupd_metadata_2022-02-22.feather

The data can be found in the above link. I extracted some of the data from that link for our task.

After getting the data we took only two types of patents that are accepted and rejected and discarded remaining of them.After that text data is converted into tokens using **distilbert** tokenizer .Then the tokenizers are fed into **distilbert transformer** and trained the model.Here I ran for 3 epochs because of the hardware limitations.Then we used the trained model to predict on the validation data and got an accuracy around 70%.After that we deployed the model in hugging face using streamlit.In hugging face first we created requirement.txt for installing required libraries and in the app.py file we had written our code.We also stored the weights after training them and uploaded them in the hugging face,because we use them for the prediction.Based on you patent application,abstract and claims it will be predict the patent predictability score.

Milestone 4: Documentation and Video Production

Goal : To create a google site to for landing the USPTO application and creating a demonstration video for running the application and documentation.

Link for google sites: https://sites.google.com/view/ustpo-app

Link for video :
https://drive.google.com/file/d/1We4AemIbHa_WCXpvgNZGsAzdUWE0BTQU/view?usp=sharing

In the video link I had explained the internals of code and application app deployed in the huggingface.Please go through it for thorough understanding.

Conclusion: The app will be will be used by patent applicants to determine the patentability of their inventions before they file their patent applications, therefore reducing the workload of the patent examiners.

The entire end to end project can be found in the link:
https://github.com/venkataseetharam