Build and Deploy LSTM Model with AWS Sagemaker

Business Overview

The Weather is one of the many factors affecting the sales of any product. In many retail stores, weather plays a vital role in the demand for consumable products. As a result, we need to understand the magnitude of the influence of various weather conditions. One of the Burger brands has a chain of retail stores across ten regions. They want to know whether or not any feature related to weather affects the sales of burgers or not. Their primary focus is to forecast Burger sales based on weather conditions. In this project, we would also deploy the model using AWS Sagemaker.

Aim:

1. **How to fetch data from SQL server hosted on Amazon RDS to Python**
2. **Whether weather influences Burger sales or not**
3. **Build LSTM model on AWS Sagemaker**
4. **To forecast the next day's Sales based on historical data**
5. **To deploy the model by using AWS Sagemaker**

Data Description:

The data is available for the sales of the burger brand from 1/1/2014 to 15/9/2020. There are a total of 24500 rows and eight columns. The data contains information about all the stores across ten different regions.
The data is stored in the MYSQL server hosted on amazon RDS.

Tech Stack

- ➔ Language: Python
- ➔ Libraries: pandas, numpy, tensorflow, boto3, pymysql, matplotlib, sklearn, tarfile, keras, lightgbm
- ➔ Platform: AWS Sagemaker, AWS S3

Approach

1. Loading data from the database using pymysql library
2. Exploratory Data Analysis
    a. Plotting date vs sales graph based on regions
    b. Autocorrelation and partial correlation plot

    c. Plotting average, maximum and minimum of features
    d. Scatterplot and histograms of the features
3. Data Preprocessing
4. Windowing the Data
5. Train Validation split
6. Data Rescaling using QuantileTransformer
7. Building baseline model(lightgbm) and feature importance
8. Building LSTM model
9. Predictions for one day into future
10. Creating an endpoint and deploying the model

Code

Once you unzip the code.zip, you can find the following files within it.

1. Burger_data.csv
   It contains the data used in this project.
   Note: In this project we have fetched the data from MySQL DB hosted on amazon RDS. It is assumed that the user has received or created credentials to fetch the data.

2. burger_stores.ipynb
   This is main file where the model is trained and deployed using sagemaker.

3. Config.yml
   Store the credentials required to connect to the DB into this file.

4. empty_train.py
   This empty train file is created while model deployment on sagemaker

5. Predict_from_endpoint.ipynb
   This file is used for prediction using the endpoint

6. Requirements.txt
   The requirements.txt file has all the required libraries with respective versions. Kindly install the file by using the command **pip install -r requirements.txt**

7. utils.py
   This file contains all the functions called in burger_stores.ipynb file.

Project Takeaways

1. How to setup AWS Sagemaker?
2. How to fetch data from MySQL Database using pymysql?
3. Brief introduction to Autocorrelation and Partial Correlation
4. How to build lightgbm model?
5. Brief introduction to LSTM
6. How to build LSTM model?
7. What is Rescaling and why it is important?
8. Rescaling data using QuantileTransformer
9. Forecasting next days sales using LSTM model
10. How to load models on AWS S3 using boto3?
11. How to create an endpoint and serve it using AWS Sagemaker?
12. How to deploy model using Sagemaker?
13. Predictions using endpoint