THE FUTURE SALES PREDICTION

PROGRESS USING PYTHON

Phase 3 submission Documents

Project Title : The future sales prediction

Analysis

Name: SOWMIYA.D

Reg No:712221205037

College: park college of engineering and technology

Sales forecasting

It is determining present-day or future sales using data like past sales, seasonality, festivities, economic conditions, etc.

So, this model will predict sales on a certain day after being provided with a certain set of inputs.

In this model 6 parameters were used as input:

- Past seven day sales
- Day of the week
- Date – the date was transformed into 3 different inputs
- Season
- Festival or not
- Sales on the same day in the previous year

Required packages and Installation

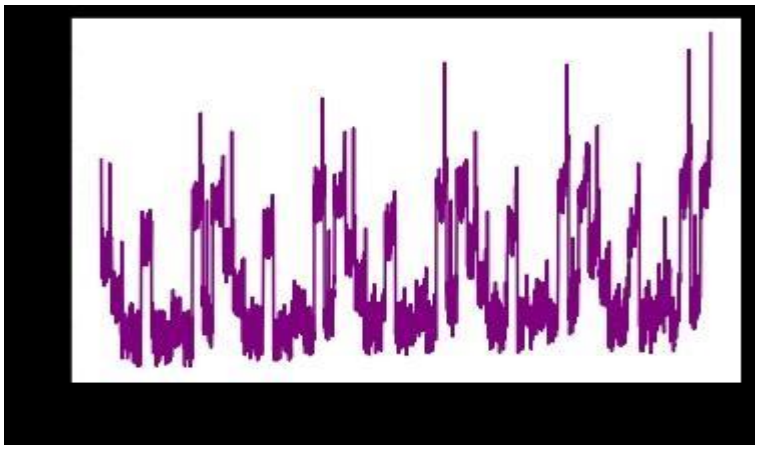
- Numpy
- Pandas
- Keras
- Tensorflow
- Csv
- Matplotlib.pyplot

```
import pandas as pd                          # to extract data from dataset(.csv file)
import csv                                   #used to read and write to csv files
import numpy as np                           #used to convert input into numpy arrays to
import matplotlib.pyplot as plt              #to plot/visualize sales data and sales fore
import tensorflow as tf                      # acts as the framework upon which this mode
from tensorflow import keras                 #defines layers and functions in the model

#here the csv file has been copied into three lists to allow better availability
list_row,date,traffic = get_data('/home/abh/Documents/Python/Untitled Folder/Sales_d
```

Original data set for sales data for 5 years:

Sales data from Jan 2015 to dec 2019



In this 5-year time frame, the first 4 years will be used to train the model and the last year will be used as a test set.

Now, a few helper functions were used for processing the dataset and creating inputs of the required shape and size. They are as follows:

1. get_data – used to load the data set using a path to its location.
2. date_to_day – provides a day to each day. For example — 2/2/16 is Saturday and 9/5/15 is Monday.
3. date_to_enc – Encodes data into one-hot vectors, this provides a better learning opportunity for the model

.

Preprocessing:

Initially, the data set had only two columns: date and traffic(sales).

After the addition of different columns and processing/normalization of values, the data contained all these values.

- Date
- Traffic
- Holiday or not
- Day

Each of them is a NumPy array of length 5 with 1s and 0s denoting its value

- We will now process some other inputs that were remaining, the reason behind using all these parameters is to increase the efficiency of the model, you can experiment with removing or adding some inputs.

- Sales data of the past seven days were passed as an input to create a trend in sales data, this will the predicted value will not be completely random similarly, sales data of the same day in the previous year was also provided.

- The following function(other_inputs) processes three inputs:

Sales data of past seven days

- Sales data on the same date in the previous year
- Seasonality – seasonality was added to mark trends like summer sales, etc.