



Final Project Report

Supervisor : Dr. Hadi Farahani

Teacher Assistant : Ali Sharifi

Prepared by Farzaneh Ahmadi (400422012)

Data mining course - Shahid Beheshti University

#All codes are written in google colab in python, (the code is provided in attachment)

Task 1: Time series interpolation service

In this task we should generate a service that its goal is to fill missing values in a time series. The input is a json file with this format:'

```
{
  "data": {
    "1400/01/01": "100",
    "1400/01/02": "150",
    "1400/01/03": "200",
    "1400/01/22": "800"
  },
  "config": {
    "type": "shamsi",
    "time": "daily",
    "interpolation": "linear"
  }
}
```

1. Install dependencies

At first we import packages and install dependencies.

2. Upload Json file

We use the code below to upload and store the Json file. `uploaded = files.upload()`

3. Read each data and config from uploaded Json separately into Pandas Dataframes

Since uploaded file is in the Python dictionary format (with key as name of uploaded file and corresponding value as the contents of the file), we change the format and decode it into "utf-8".

Each data will be read and configured from json separately into pandas dataframes. Sample:

	Date	price
0	1400/01/01	100
1	1400/01/02	150
2	1400/01/03	200
3	1400/01/22	800
		option
	params	
	type	shamsi
	time	daily
	interpolation	linear

4. Convert the Shamsi date to Gregorian

To simplify the process we convert the Shamsi date to Gregorian (If needed). By using a proper function, the data will be converted to Gregorian. At the end after doing the analysis, we will convert data to Shamsi again. The example data format given above will be turned into the Gregorian format as shown in the picture below:

	price
Date	
2021-03-21	100
2021-03-22	150
2021-03-23	200
2021-04-11	800

5. Generate the missing dates

We generate missing dates based on defined frequency. Missing dates between the first and the last dates will be generated with a NaN value.

2021-03-22	150.0
2021-03-23	200.0
2021-03-24	NaN
2021-03-25	NaN
2021-03-26	NaN
2021-03-27	NaN
2021-03-28	NaN
2021-03-29	NaN
2021-03-30	NaN
2021-03-31	NaN

6. Interpolate missing values

The missing values will be estimated according to the type of interpolation (linear, spline, polynomial). In the example in this report, we consider linear interpolation.

linear	
Date	price
2021-03-21	100.000000
2021-03-22	150.000000
2021-03-23	200.000000
2021-03-24	231.578947
2021-03-25	263.157895
2021-03-26	294.736842
2021-03-27	326.315789
2021-03-28	357.894737
2021-03-29	389.473684
2021-03-30	421.052632
2021-03-31	452.631579
2021-04-01	484.210526
2021-04-02	515.789474
2021-04-03	547.368421

7. Return the Gregorian to Shamsi date

In step 4, we converted Shamsi dates to Gregorian. We now return it back to Shamsi. By using a proper function (i.e., `to_jalali_date`), time series data will be turned into the Shamsi format.

8. print the output Json

```
{'1400/01/01': 100.0,  
'1400/01/02': 150.0,  
'1400/01/03': 200.0,  
'1400/01/04': 231.57894736842104,  
'1400/01/05': 263.1578947368421,  
'1400/01/06': 294.7368421052631,  
'1400/01/07': 326.3157894736842,  
'1400/01/08': 357.89473684210526,  
'1400/01/09': 389.4736842105263,  
'1400/01/10': 421.0526315789474,  
'1400/01/11': 452.63157894736844,  
'1400/01/12': 484.2105263157895,  
'1400/01/13': 515.7894736842105,  
'1400/01/14': 547.3684210526316,  
'1400/01/15': 578.9473684210526,
```

9. Writing data into a file

We can also write the result into a Json file using code below

```
#we can also write the result into a Json file  
dictionary = ts.set_index('Date')['price'].to_dict()  
with open('json_data.json', 'w') as outfile:  
    json.dump(dictionary, outfile)
```

Source Code:

<https://colab.research.google.com/drive/16ww51d12oAMywNBjn-oEIDLlBnMX-ypB?usp=sharing>

Task 2: Time series interpolation service

This task is same as task1. But in this task we should convert the data to Shamsi and then doing interpolation. The input is a json file with this format:'

```
{
  "data": {
    "2022/01/01": "100",
    "2022/01/02": "150",
    "2022/01/03": "200",
    "2022/01/22": "800"
  },
  "config": {
    "type": "miladi",
    "time": "daily",
    "interpolation": "linear"
  }
}
```

This service works same as the previous service, but just they differ in step 4 and 7. In this service in the step 4, the data will converted from Gregorian to Shamsi and then interpolation will be done. and step 7 will be deleted and the output is in Shamsi format.

Source Code:

https://colab.research.google.com/drive/1racqNM2qW_F8kXyCo9iJVJnAx1-dZtUv?usp=sharing