پروژه نهایی داده کاوی

محمد زیاری - 97222047

سرویس 1 : در این سرویس ورودی به شکل زیر است :

```
{
{ :"data"
{ :"time"
"0": "1999-2-3"
"1":"1999-2-5"
"2":"1999-2-6"
},
{ :"vol"
"0": 20 ,
"1": 40 ,
"2": 100 ,
}
,}
{ :"config"
"type": "miladi" or "shamsi"
,"time": "monthly" or "daily"
,"interpolation":"pad" or "linear" or "polynomial" or "spline"
"order": int
}
}
```

- ماه و سال و روز را با - جدا میکنیم.
- قابل ذکر است وقتی interpolation را polynomial یا spline میگذاریم order را می توانیم عددی وارد کنیم ولی در pad و linear این مقدار تاثیری ندارد
- خروجی در این سرویس timestamp برحسب سال میلادی و vol خواهد بود.

سرویس 2 : در این سرویس ورودی به شکل زیر است :

```
{
{ :"data"
{ :"time"
"0": "1999-2-3"
"1":"1999-2-5"
"2":"1999-2-6"
},
{ :"vol"
"0": 20 ,
"1": 40 ,
"2": 100 ,
}
,}
{ :"config"
"time": "monthly" or "daily"
,"interpolation":"pad" or "linear" or "polynomial" or "spline",
"order": int
}
}
```

- علاوه بر موارد گفته شده در سرویس 1 باید گفت که خروجی این سرویس نیز تاریخ شمسی و vol خواهد بود.

سرویس 3 : ورودی این سرویس دقیقا مشابه ورودی اولیه داده شده است و اگر بخواهیم time series بدهیم باید در config هم time series را true کنیم.


```
}
} :"data"
} :"time"
,"1990-11-14" :"0"
,"1990-11-15" :"1"
,"1990-11-16" :"2"
"1990-11-17" :"3"
,{
} :"feature"
,20 :"0"
,40000 :"1"
,70 :"2"
100 :"3"
{
,{
} :"config"
time_series": true"
{
{
```


- دو روش برای داده های time series نیزیکی استفاده از پکیج prophet و فیت کردن مدل روی آن و مقایسه با حد upper و lower بدست آمده از y hat است که متغیری است که خود پکیج prophet به ما می دهد. روش دوم نیز استفاده از autoregression بود که به نظر prophet نتایج بهتری داشت و فکر می کنم شاید autoregression روی داده های بزرگتر بهتر نتیجه می دهد.
- دو روش برای داده های غیر time series نیز استفاده از چارک های اول و سوم و فواصل 1.5 برابر کمتر و بیشتر از آنهاست و دیگری استفاده از روش قدیمی بین 3 برابر کمتر یا بیشتر فاصله از std نسبت به mean می باشد. در این روش ها هم چارک روی داده های کوچک بهتر از روش دوم جواب می دهد.

سرویس 4 : ورودی این سرویس مطابق چیزی که وجود داشت

```
}
} :"data"
} :"id"
,1 :"0"
,2 :"1"
,3 :"2"
,4 :"3"
,5 :"4"
6 :"5"
,{
} :"feature1"
,50 :"0"
,12 :"1"
,50 :"2"
,500 :"3"
,60 :"4"
12 :"5"
,{
} :"class"
,1 :"0"
,1 :"1"
,1 :"2"
,1 :"3"
,1 :"4"
0 :"5"
{
,{
} :"config"
,major_class": 1"
,minor_class": 0"
method": "SMOTE" or "undersampling " or "oversampling " or "Clustercentroids"
" or "Tomeklinks
{
{
```

- متدهای موجود smote و oversampling و undersampling هستند که اجباری بود. علاوه بر آن دو متد Clustercentroids و Tomeklinks هستند که همه با پکیج پیاده سازی شده اند.
- بقیه چیزها دقیقا مشابه data و config تعریف شده می باشند و خروجی به شکل خواسته شده در سوال می باشد.

نکات مهم

- برای دسترسی به سرویس ها نیز باید روی localhost:5000/service1 انجام شود. برای سرویس های بعدی نیز جای 1 عددی دیگر می گذاریم.
- فایل requirements هم در پروژه موجود است.
- ورودی و input در فایل پروژه نمونه ای آپدیت شده و درست است ولی خروجی آپدیت نشده است.