商务智能第一次作业

何子昂 **2020302111281**

Definition of Business Intelligence

- Broad Definition: An umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies
- Narrow Definition: Descriptive analytics tools and techniques

The legal problem in Business Intelligence

Crime of infringing personal information of citizens

The crime of infringing personal information of citizens is very common in Business Intelligence cases especially those adopting crawlers to acquire the information of users.

According to the article provided by the Supreme People's Procuratorate of the People's Republic of China, some cases are listed:

> *Hangzhou Moscorpion Data Technology Co., Ltd. explicitly informs the user in the agreement that the company will not save the user's account password and other information, but without the user's permission, the company still uses crawler code technology to save more than 20 million users' various accounts and passwords on its rented server for a long time. Through secondary processing, the products are provided to online loan companies as "risk control", and fees ranging from 0.1 yuan to 0.3 yuan are charged from online loan platforms for each transaction.*

We can find that even if the company has told users its storage of their information, it exceeded its power and utilized them to make profits.

Business Analytics

Descriptive Analytics

- Descriptive or reporting analytics what is happening in the organization and understanding some underlying trends and causes of such occurrences.
- Answering the question of what happened/ what is happening

Predictive Analytics

- Aims to determine what is likely to happen in the future (foreseeing the future events)
- Answering the question of what will happen and why will it happen

Prescriptive Analytics

- Aims to determine the best possible decision
- Answering the question of what should I do and why should I do it
- Create the alternatives, and then determines the best one

Application Case

> *Organization: Des Moines Public Schools*
>
> *Problem: Manual Excel reporting means administrators don't have access to up-to-date data (such as attendance) and therefore can't intervene in a timely manner.*
>
> *The solution: Des Moines Public Schools (DMPS) uses advanced analytics to improve dropout intervention rates, as well as to better understand the impact of various teaching methods on individual student achievement.*
>
> *The DMPS research and data management team uses a multiple linear regression model, commonly known as the dropout coefficient, to assign weights to various student indicators to predict which students may be at risk of dropping out. They used a business intelligence platform to leverage the model. Through data visualization, employees can easily identify individual students at risk and get them the attention they need.*
>
> *The dashboard set up by the research and data management team provides real-time analysis to 7,000 DMPS teachers and staff, allowing them to make adjustments and interventions faster, resulting in significantly improved intervention success rates. They can view five years of historical data in the real-time analysis. This means that staff can delve into historical data on site, as a way to validate insights that are relevant to current students.*

In this case, BI is typically applied. By using advanced analytics, DMPS can easily gain the current situation of the students study in the school, which is called descriptive analytics. And with the help of the multiple linear regression model, they will be able to predict which students may be at risk of dropping out. And this is the predictive analytics. After knowing the group at the risk of dropping out, diverse solutions to help them out gained through BI are the implements of prescriptive analytics.

Skewness and Kurtosis

Skewness

Measure of asymmetry

$$Skewness = S = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)s^3}$$

Example

- $\mu = 1, \sigma = 1 \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-1)^2}{2}) : S = 0$
- $\mu = -1, \sigma = 1 \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{(x+1)^2}{2}) : S = 0$

Kurtosis

Peak/tall/skinny nature of the distribution

$$Kurtosis = K = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4} - 3$$

Example

- $\mu = 1, \sigma = 1 \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-1)^2}{2}) : K = 0$
- $\mu = -1, \sigma = 1 \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{(x+1)^2}{2}) : K = 0$

for gaussian distribution, its skewness and Kurtosis are both nearly to 0.

Dissimilarity Matrix

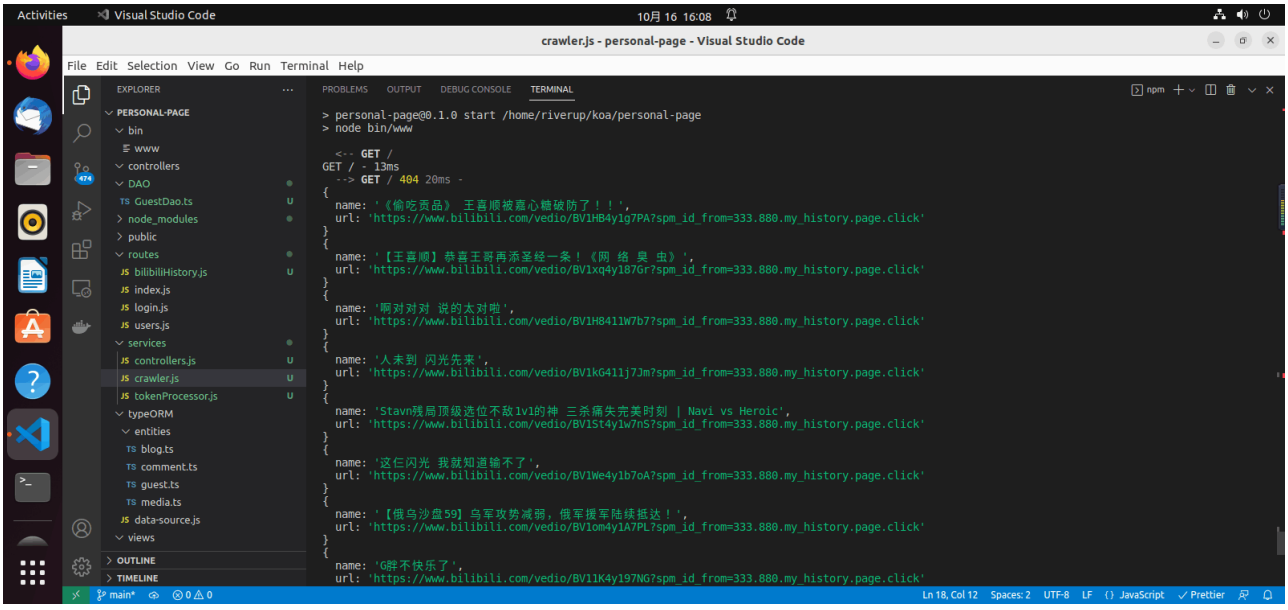
Data

NAME	DEGREE	PLAY BADMINTON OR NOT	PLAY BASKETBALL OR NOT	AGE	LOCATION	RELATION
Me	bachelor	P	P	20	Wuhan	Me
Mother	graduate	N	N	46	Benxi	Related
Luo	master	P	N	22	Wuhan	Friend
Huo	bachelor	N	P	20	Peking	Friend
Zheng	bachelor	N	N	19	Dongjing	Friend
Qu	bachelor	N	P	19	Peking	Friend

Matrix

1					
0.99	1				
0.43	0.73	1			
0.50	0.82	0.60	1		
0.68	0.67	0.43	0.34	1	
0.51	0.83	0.60	0.01	0.33	1

Crawler



```
const superagent = require("superagent");//use the superagent to
crawl the api
cookies_dict =
  "buvid3=26FE2565-FBA9-70DA-9BAD-4494DC96128237091infoc; rpdid=|
(u))u~k||mlOJ'uYu|lmYJRu; _uuid=335F55EB-FB710-8BA1-B4E5-
63F10A9548AC122293infoc; i-wanna-go-back=-1;
LIVE_BUVID=AUTO8916429522535161; buvid4=D98E8518-4C94-2C58-ED79-
D8FAFC5F62F994360-022012423-OOXTsSUsuhXKPhilOAzTTg%3D%3D;
CURRENT_BLACKGAP=0; nostalgia_conf=-1; hit-dyn-v2=1;
fingerprint3=491aac715399b61a0ad45ad9565fbea6;
fingerprint=a26cbb3ab8190ee35b39adcf0adf5775;
buvid_fp_plain=undefined; DedeUserID=673457443;
DedeUserID__ckMd5=f0c766782e7db7f6;
buvid_fp=a26cbb3ab8190ee35b39adcf0adf5775; b_ut=5;
blackside_state=0; is-2022-channel=1; b_nut=100;
CURRENT_QUALITY=112; PVID=1; innersign=0;
b_lsid=A41B16C5_183DC6239AE;
SESSDATA=9356ce69%2C1681401770%2Cb94cc%2Aa1;
bili_jct=ebf5c01815845a1c034a5d9d584fdf96; CURRENT_FNVAL=16;
sid=8e7qqzf9;
bp_video_offset_673457443=717306929269964900";//register the cookie
function crawlBiliBili() {
  var requestHistories = new Array();
  superagent
    .get("https://api.bilibili.com/x/web-
interface/history/cursor")//visit the api
    .set("Cookie", cookies_dict)//with the header of cookie
    .end(function (err, res) {
      if (err) {
        console.error();
      } else {
        var str = res.text;

          /*match the useful information*/
        str = str.match(/(?<="list":\[).*(?=]}})/)[0];
        var strs = str.split(/(?<=}),(?={)/);
        strs.forEach((element) => {

            /*parse the string to json object*/
          requestHistories.push(JSON.parse(element));
        });
        var histories = new Array();
        requestHistories.forEach((element) => {
```

```
            var history = new Object();
            history.name = element.title;
            history.url =
`https://www.bilibili.com/vedio/${element.history.bvid}?
spm_id_from=333.880.my_history.page.click`;
            histories.push(history);
          });
          histories.forEach((element) => {


            console.log(element);
          });
        }
      });
}
```