بسم الله الرحمن الرحيم

سرطان روده بزرگ (Colon Cancer)

پروژه درس داده کاوی

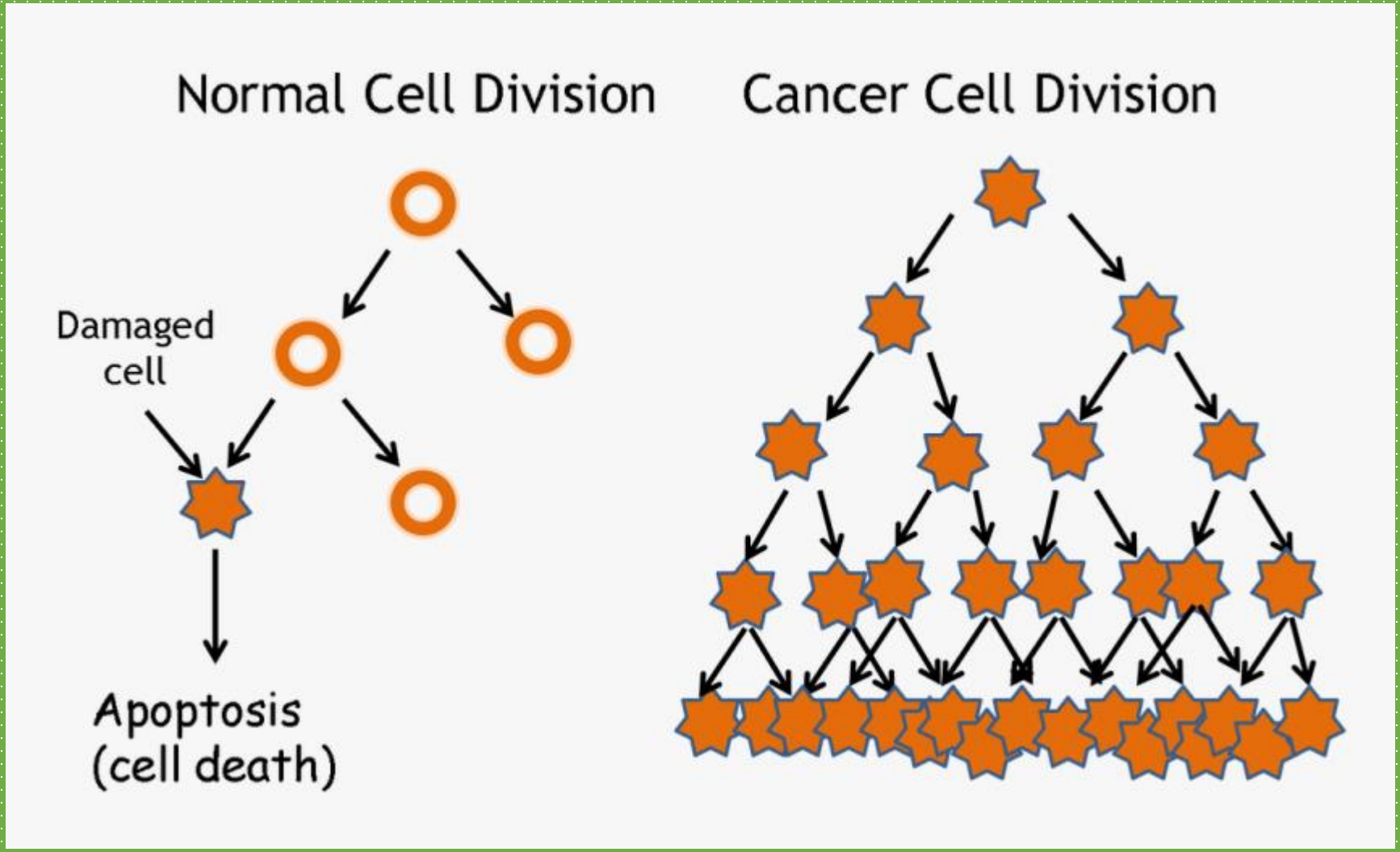
استاد درس: دکتر الهام عباسی

اسفند ۹۹

فهرست


معرفی مسأله
داده
data
پیش‌پردازش
preprocessing
مدل
Model
ارزیابی
Evaluation

سرطان چیست؟


Normal Cell Division
Cancer Cell Division
Damaged cell
Apoptosis (cell death)

سرطان روده بزرگ

TABLE 1. New Cases and Deaths for 36 Cancers and All Cancers Combined in 2020

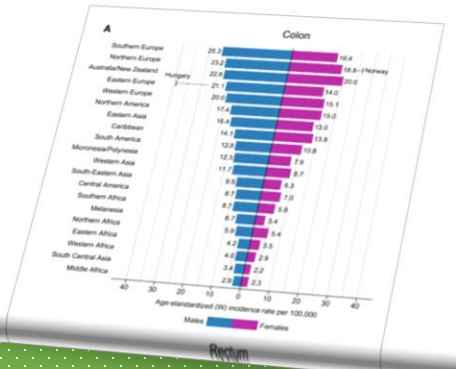
CANCER SITE	NO. OF NEW CASES (% OF ALL SITES)		NO. OF NEW DEATHS (% OF ALL SITES)	
Female breast	2,261,419	(11.7)	684,996	(6.9)
Lung	2,206,771	(11.4)	1,796,144	(18.0)
Prostate	1,414,259	(7.3)	375,304	(3.8)
Nonmelanoma of skin [a]	1,198,073	(6.2)	63,731	(0.6)
Colon	1,148,515	(6.0)	576,858	(5.8)
Stomach	1,089,103	(5.6)	768,793	(7.7)
Liver	905,677	(4.7)	830,180	(8.3)
Rectum	732,210	(3.8)	339,022	(3.4)
Cervix uteri	604,127	(3.1)	341,831	(3.4)
Esophagus	604,100	(3.1)	544,076	(5.5)
Thyroid	586,202	(3.0)	43,646	(0.4)
Bladder	573,278	(3.0)	212,536	(2.1)
Non-Hodgkin lymphoma	544,352	(2.8)	259,793	(2.6)
Pancreas	495,773	(2.6)	466,003	(4.7)
Leukemia	474.519	(2.5)	311.594	(3.1)
All sites excluding nonmelanoma skin	18,094,716		9,894,402	
All sites	19,292,789		9,958,133	

...rectal cancer can be considered a marker of socioeconomic development, and, in ...countries undergoing major transition, incidence rates tend to rise uniformly with increasing HDI.[92, 93] Incidence rates have been steadily rising in many countries in Eastern Europe, South Eastern and South Central Asia, and South America.[22, 94] The increase in formerly low-risk and lower HDI countries likely reflects changes in lifestyle factors and diet, ie, shifts toward an increased intake of animal-source foods and a more sedentary lifestyle, leading to decreased physical activity and increased prevalence of excess body weight, which are independently associated with colorectal cancer risk.[95] Additional risk factors include heavy alcohol consumption, cigarette smoking, and consumption of red or processed meat, whereas calcium supplements and adequate consumption of whole grains ... products appear to decrease risk.[96] Primary prevention remains ... the increasing global burden of colorectal cancer. The expens... screening effort in most LMICs is not currently justified giver... colonoscopy and inadequate implementation of diagnostic... evidence, however, suggests that colorectal cancer screeni... invasive methods (guaiac testing and fecal immunochemic... least in some settings of emerging economies, and offer o... burden of the disease.[97, 98]

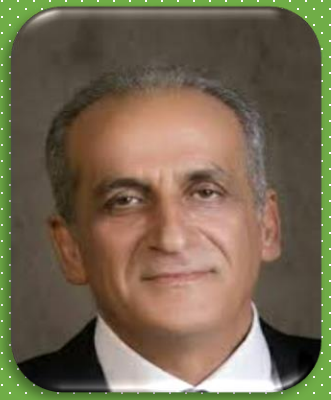
Declines in colorectal cancer incidence in some high-incid... attributed to population-level changes toward healthier l... smoking) and the uptake of screening,[94, 99] although acc... 2000s is chiefly attributed to increased colonoscopy scre... lesions.[100-102] However, favorable trends for adults age... early-onset colorectal cancer (age at diagnosis <50 year... ...nited States, Canada, Australia, and 6 other high-inco... ...4% per year.[103-105] Although rising incidence in y...

Colorectal cancer

More than 1.9 million new colorectal cancer (including anus) cases and 935,000 deaths were estimated to occur in 2020, representing about one in 10 cancer cases and deaths (Table 1). Overall, colorectal ranks third in terms of incidence, but second in terms of mortality (Fig. 4). Incidence rates are approximately 4-fold higher in transitioned countries compared with transitioning countries, but there is less variation in the mortality rates because of higher fatality in transitioning countries (Fig. 7). There is an approximately 9-fold variation in colon cancer incidence rates by world regions, with the highest rates in European regions, Australia/New Zealand, and Northern America, with Hungary and Norway ranking first in men and women, respectively (Fig. 10A). Rectal cancer incidence rates have a similar regional distribution, although rates in Eastern Asia rank among the highest (Fig. 10B). Rates of both colon and rectal cancer incidence tend to be low in most regions of Africa and in South Central Asia.



دانش ما و سرطان



مکتب‌خونه

ریاضیات سرطان

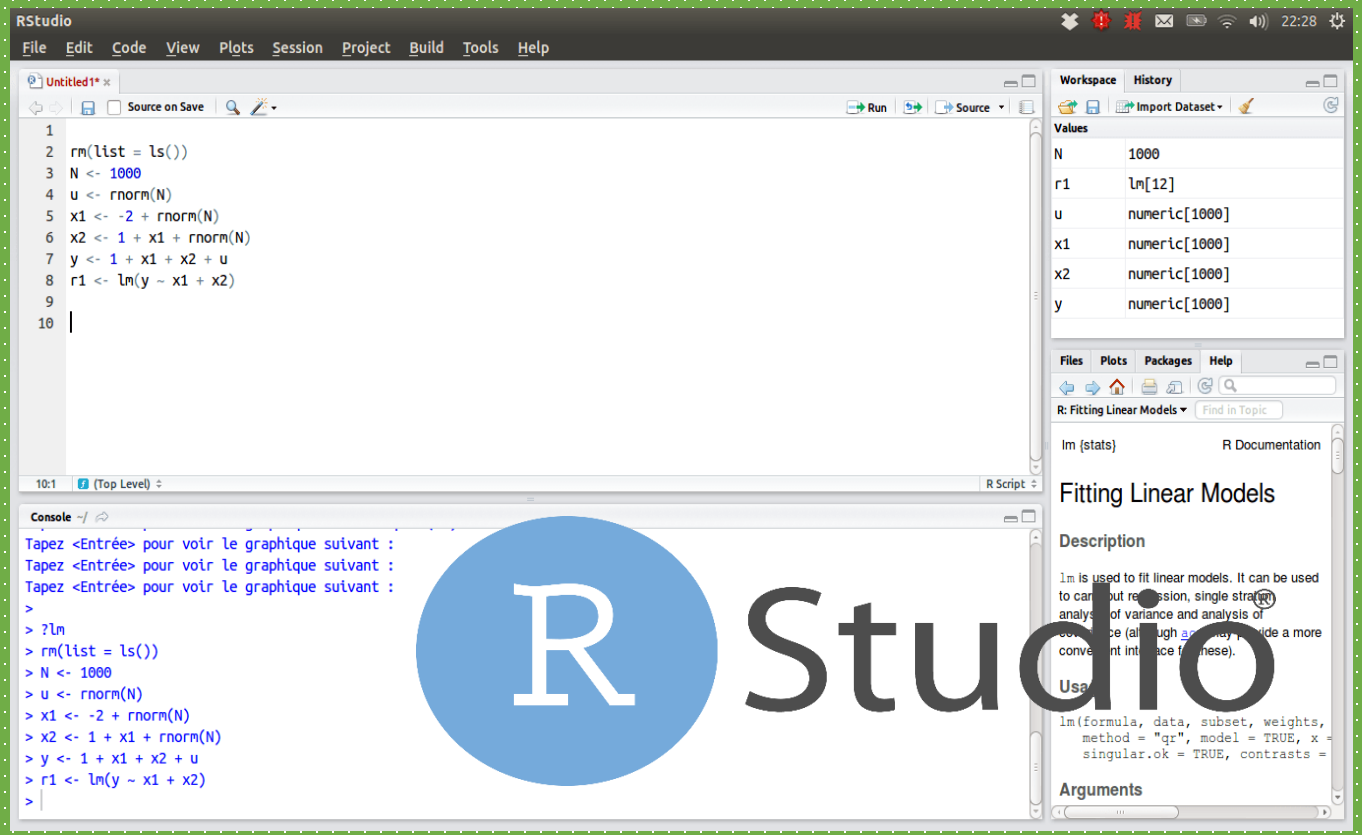
۸۲٪ (۲۹ رای)

ثبت‌نام رایگان

نمایش تیزر

زبان برنامه نویسی R

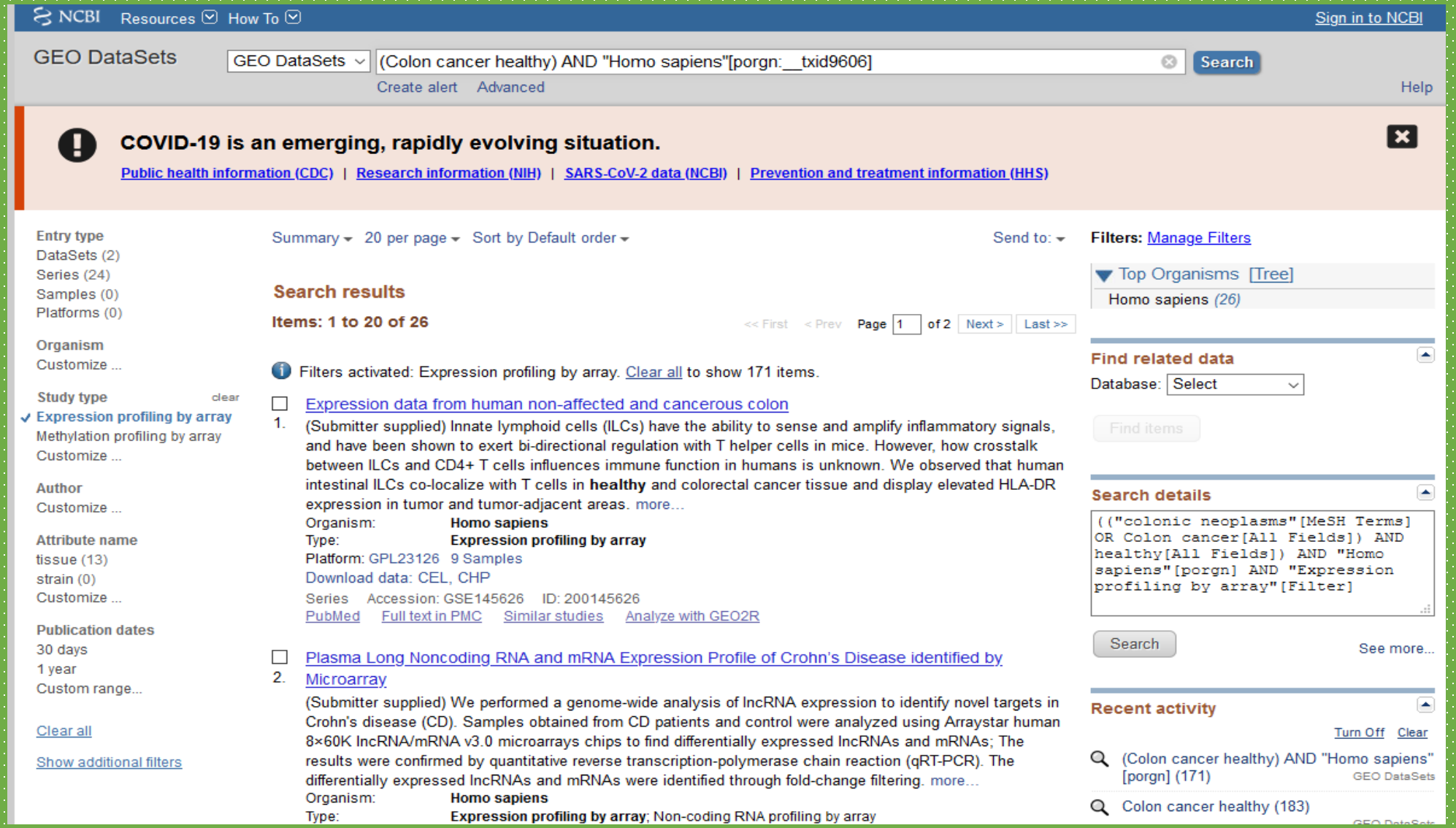
- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

National Center for Biotechnology Information (NCBI)

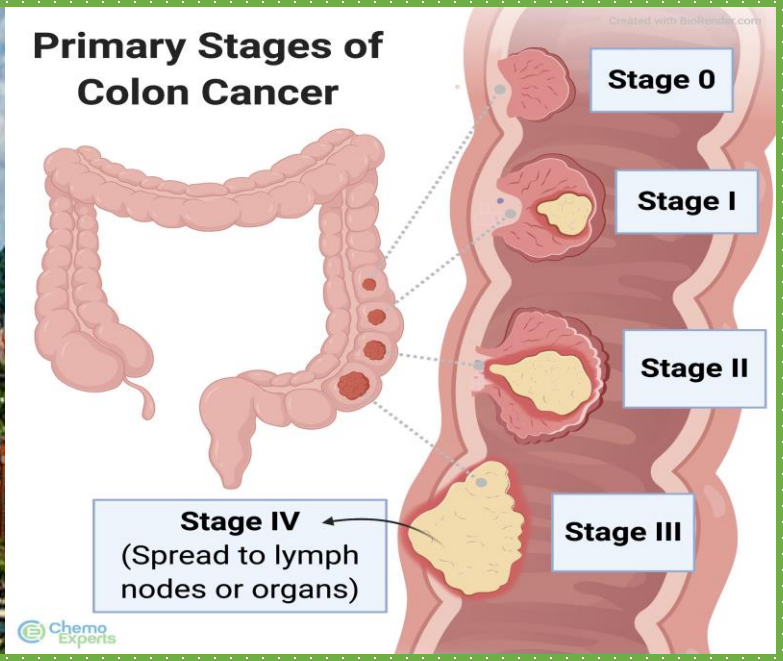
خواندن داده



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

National Taiwan University

خواندن داده



colon-cancer

- Source: [AU99a]
- Preprocessing: Instance-wise normalization to mean zero and variance one. Then feature-wise normalization to mean zero and variance one. [SKS03a]
- # of classes: 2
- # of data: 62
- # of features: 2,000
- Files:
  - colon-cancer.bz2

Cancer ~ 2000 Gene Expression

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

پیش پردازش

داده های نرمال و هم توزیع شده

C:\Users\shamsollah\Downloads\Data Mining\Project\colon-cancer - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

change.log | web.xml | context.xml | DK_predict.py | KNN.py | SecondaryStruct.py | DecisionTree.py | PebblesGame.java | ColonCancer.R | note 1.Rmd | colon-cancer

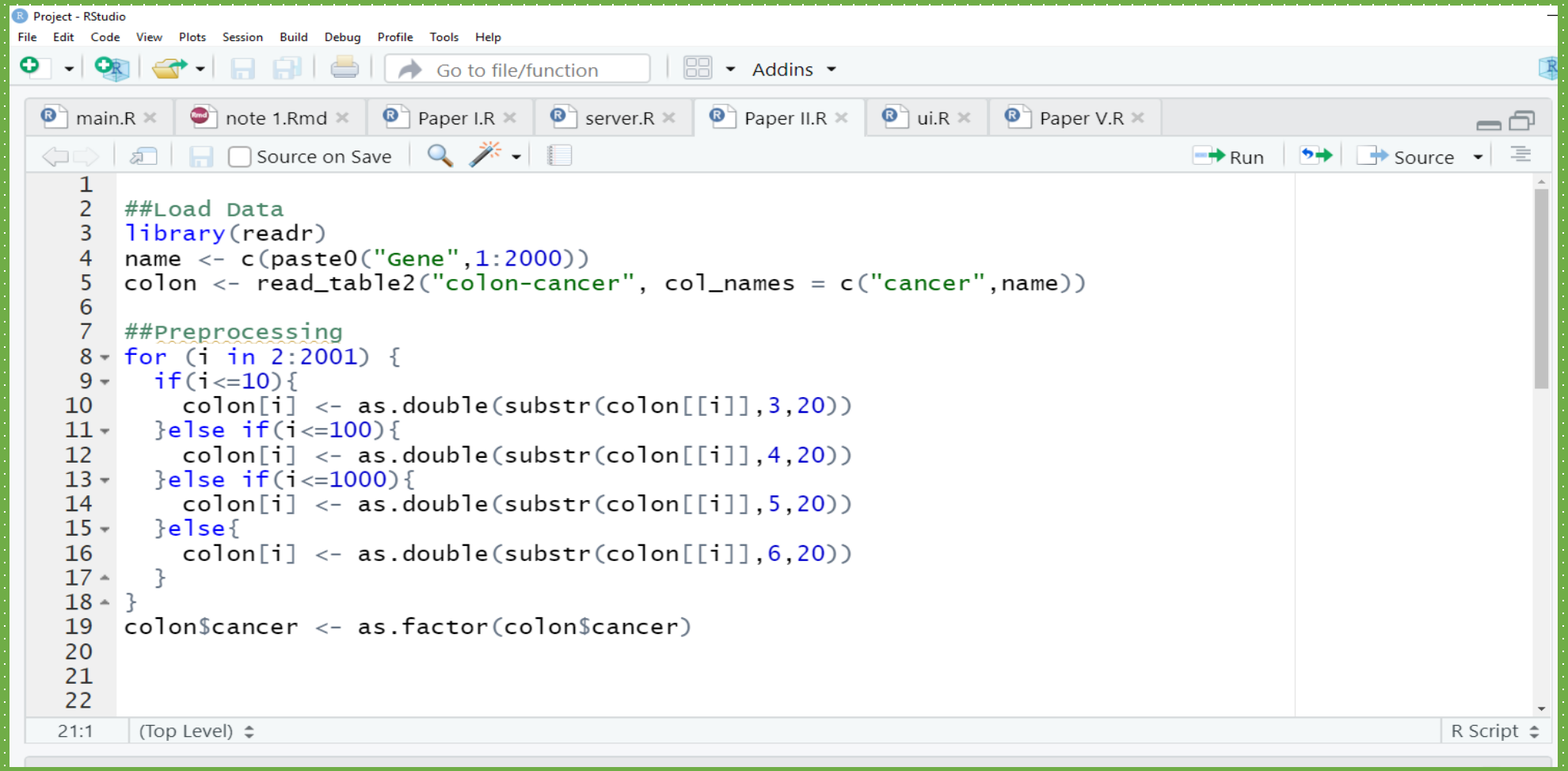
```
-1.000000  1:2.080750 2:1.099070 3:0.927763 4:1.029080 5:-0.130763 6:1.265460 7:-0.436286 8:0.728881 9:2.107980 10:1.359870 11:0.265471 12:-0.324436 13:-1.200930 14:-1.531920 15:0.298026 16:0
1.000000  1:1.109460 2:0.786453 3:0.445560 4:-0.146323 5:-0.996316 6:0.555759 7:0.290734 8:-0.145259 9:1.132660 10:0.559093 11:-1.469130 12:-0.813699 13:0.594290 14:0.404811 15:-1.197510 16:0
-1.000000  1:-0.676530 2:1.693100 3:1.559250 4:1.559980 5:-0.982179 6:-1.358510 7:-1.313990 8:-0.455067 9:0.295214 10:0.290694 11:0.415632 12:0.136703 13:-0.688848 14:-1.700350 15:-1.864170 1
1.000000  1:0.534396 2:1.677540 3:1.489030 4:0.778605 5:-0.183776 6:-1.116850 7:-1.487560 8:-0.579511 9:0.292683 10:1.345480 11:-0.687898 12:-0.749802 13:-1.168640 14:-0.559310 15:-0.818024 1
-1.000000  1:-1.018900 2:0.511080 3:0.755641 4:1.013820 5:0.529899 6:0.160440 7:-0.087055 8:1.295290 9:0.458736 10:0.714082 11:0.727290 12:1.558500 13:1.340430 14:0.486905 15:0.876190 16:0.5
1.000000  1:-1.185370 2:-0.514473 3:-0.566634 4:1.224720 5:0.619244 6:-0.684713 7:-0.798129 8:1.368770 9:-0.697007 10:-1.006190 11:0.808748 12:1.677320 13:0.847126 14:1.930790 15:0.944730 16:
-1.000000  1:1.779050 2:0.423947 3:0.820696 4:2.525690 5:0.666921 6:0.661346 7:0.425365 8:0.165247 9:1.967910 10:1.140080 11:-0.779230 12:0.074396 13:-0.660087 14:-0.976049 15:0.159321 16:0.2
1.000000  1:-0.889638 2:-0.315453 3:-0.073131 4:1.157500 5:-0.311039 6:-0.364472 7:-1.621640 8:1.193000 9:0.689805 10:0.203786 11:1.907460 12:0.734614 13:0.378114 14:1.156440 15:-0.635016 16:
-1.000000  1:-0.659694 2:-0.184388 3:-0.540022 4:1.122420 5:0.562609 6:-2.988310 7:-2.349810 8:-1.325010 9:-0.017002 10:0.109081 11:1.005260 12:-1.120500 13:-2.166720 14:-0.588374 15:0.475411
1.000000  1:-1.225800 2:-0.212615 3:-0.588923 4:1.335410 5:-0.356505 6:0.354394 7:0.699607 8:0.190782 9:-0.139117 10:-0.551846 11:0.292320 12:0.596870 13:2.388360 14:1.380680 15:-0.303903 16:
-1.000000  1:-0.377282 2:-2.620490 3:-2.763910 4:0.612038 5:-0.155718 6:-1.456070 7:0.683292 8:-1.031570 9:0.718861 10:-2.313070 11:-0.092949 12:-0.624663 13:0.459021 14:-1.306700 15:-0.3568
1.000000  1:1.639210 2:0.663551 3:0.324924 4:1.456460 5:-0.170587 6:-0.443486 7:-0.832590 8:-0.928334 9:1.320070 10:0.591865 11:0.158091 12:-0.703188 13:0.562021 14:1.029230 15:-0.092680 16:-
-1.000000  1:0.125390 2:1.286660 3:1.505230 4:0.804243 5:1.384430 6:0.633277 7:1.682130 8:1.925000 9:0.072503 10:0.328855 11:0.792426 12:1.454890 13:0.790981 14:0.684712 15:0.879764 16:0.7785
1.000000  1:0.055593 2:0.758397 3:0.857047 4:0.159382 5:1.020290 6:0.733855 7:1.141090 8:1.213820 9:-0.236974 10:0.442875 11:0.872624 12:0.797200 13:0.381540 14:1.014760 15:0.683497 16:1.3595
-1.000000  1:-0.861022 2:0.432429 3:0.404421 4:-0.230533 5:0.492867 6:0.274014 7:-0.873134 8:0.609713 9:-0.112962 10:0.578550 11:1.313170 12:0.546942 13:-0.755124 14:-0.178935 15:0.591529 16:
1.000000  1:-0.672304 2:-0.484720 3:-0.667529 4:-0.703183 5:1.872860 6:1.295700 7:1.044660 8:-1.255300 9:-0.686953 10:1.347860 11:1.222190 12:-1.240440 13:-0.533714 14:-0.101042 15:0.923564 1
-1.000000  1:1.605150 2:-0.771553 3:-0.819771 4:0.292317 5:1.648190 6:0.192798 7:1.348880 8:1.284650 9:1.989650 10:-0.786998 11:2.258840 12:1.745370 13:-0.146794 14:-0.633022 15:2.003400 16:-
1.000000  1:0.867577 2:0.501470 3:0.242006 4:0.417230 5:1.723880 6:-1.249270 7:-0.815193 8:2.015220 9:1.160140 10:-0.682794 11:0.322753 12:2.374360 13:1.202120 14:1.028740 15:0.476317 16:1.88
-1.000000  1:0.590751 2:0.440645 3:0.585110 4:1.470660 5:0.724532 6:-1.874140 7:-2.038920 8:1.115610 9:1.225130 10:0.217863 11:0.772405 12:1.571610 13:-0.849927 14:0.255625 15:1.191100 16:-0.
1.000000  1:-0.311476 2:-1.482610 3:-1.697240 4:1.728990 5:0.511898 6:-0.943735 7:-0.285187 8:1.028790 9:-0.365555 10:-1.147750 11:0.551768 12:1.604300 13:1.871430 14:2.088650 15:0.929057 16:
-1.000000  1:0.369171 2:0.742324 3:0.584255 4:0.617444 5:-1.513950 6:1.180220 7:-0.268009 8:0.513435 9:0.428941 10:2.254840 11:-0.694395 12:-0.445303 13:-0.403973 14:-1.845930 15:-1.022960 16
1.000000  1:0.137575 2:0.801932 3:0.673334 4:0.089943 5:-1.572560 6:0.878675 7:0.506696 8:0.222351 9:1.146440 10:0.554650 11:-0.780461 12:-0.756995 13:1.511520 14:-0.660579 15:-1.378220 16:2.
-1.000000  1:-1.163370 2:-1.214550 3:-1.359720 4:-1.237730 5:1.509940 6:0.462035 7:0.989522 8:1.017890 9:-1.057710 10:-0.418986 11:2.009340 12:1.083270 13:0.340543 14:0.217414 15:1.903150 16:
1.000000  1:0.907830 2:1.640850 3:1.416460 4:0.953309 5:2.895000 6:0.322749 7:1.517180 8:2.689920 9:0.509099 10:0.430893 11:1.190580 12:2.987450 13:2.935570 14:1.834640 15:1.668150 16:-0.5315
-1.000000  1:0.261661 2:-0.126443 3:0.084908 4:-0.461416 5:0.420518 6:0.386768 7:0.257387 8:1.184280 9:-0.059124 10:0.305515 11:0.424351 12:1.110520 13:-0.830866 14:0.262912 15:1.554110 16:0.
-1.000000  1:1.070150 2:0.393399 3:0.695527 4:0.180488 5:-0.539207 6:0.520514 7:1.279120 8:-0.856697 9:0.544119 10:-0.196598 11:-0.931316 12:-0.786116 13:0.325911 14:-1.412390 15:-0.652764 16
-1.000000  1:1.440710 2:0.649996 3:0.873454 4:0.190425 5:-0.368266 6:1.162670 7:1.748130 8:-0.859296 9:1.142550 10:0.713684 11:-1.599660 12:-1.187890 13:0.940049 14:-1.161600 15:-0.141056 16:
-1.000000  1:0.603603 2:1.226510 3:1.338770 4:0.105289 5:0.012994 6:0.735462 7:1.281290 8:0.208414 9:0.605721 10:0.968145 11:0.540411 12:0.108103 13:-0.166173 14:-0.399909 15:0.713992 16:0.35
-1.000000  1:0.668765 2:0.444708 3:0.397101 4:0.643610 5:-1.578100 6:0.852593 7:1.579510 8:-1.791650 9:0.432917 10:0.707365 11:-1.077720 12:-1.367150 13:0.503573 14:-1.741590 15:-0.848765 16:
-1.000000  1:1.196460 2:0.259518 3:0.023820 4:0.661112 5:-0.757198 6:0.931691 7:0.942438 8:-1.034320 9:1.033310 10:0.441963 11:-1.179770 12:-0.809941 13:0.436219 14:-1.206550 15:0.058378 16:0
-1.000000  1:-0.492445 2:0.587706 3:0.682120 4:0.259068 5:0.028761 6:0.616336 7:0.271316 8:-0.376469 9:0.199924 10:2.079410 11:1.211410 12:-0.484389 13:-0.325309 14:-0.527436 15:-0.181610 16:
-1.000000  1:-0.293825 2:-2.289300 3:-2.016040 4:0.150357 5:-0.313382 6:-0.578577 7:0.625957 8:-0.703222 9:0.451249 10:-2.672270 11:0.295318 12:-0.265032 13:1.050780 14:-0.140789 15:-0.193836
-1.000000  1:0.704230 2:-1.266950 3:-1.082390 4:0.282644 5:0.599121 6:1.176300 7:0.821073 8:0.267318 9:0.377003 10:-2.000200 11:0.860445 12:0.432843 13:-0.265831 14:0.468599 15:1.240710 16:-0
-1.000000  1:0.451197 2:1.533080 3:1.493210 4:0.108655 5:0.118696 6:0.359560 7:-0.080691 8:1.088030 9:0.041763 10:0.329834 11:1.340370 12:0.972483 13:0.441282 14:0.105281 15:0.966367 16:0.023
-1.000000  1:0.173573 2:-0.086988 3:0.049312 4:0.154649 5:-0.854644 6:0.395668 7:-0.501723 8:-0.606768 9:0.427740 10:0.671266 11:0.246886 12:-0.720489 13:-0.688775 14:-0.750862 15:-0.242908 1
-1.000000  1:-0.552768 2:-0.394988 3:-0.067447 4:-1.468450 5:0.986678 6:-0.950956 7:-0.564847 8:0.906280 9:-0.336318 10:-0.852705 11:0.055996 12:0.998134 13:-0.872140 14:0.063377 15:0.771987
-1.000000  1:-1.633650 2:1.224960 3:1.036480 4:1.108610 5:-2.121200 6:0.087452 7:0.285415 8:-0.453603 9:0.292197 10:0.092917 11:-0.599566 12:0.021802 13:0.334208 14:-2.319520 15:-0.715272 16:
-1.000000  1:-1.154570 2:-1.139510 3:-0.852773 4:-1.805590 5:1.211640 6:-0.043248 7:1.020930 8:0.475199 9:-0.733753 10:-0.220390 11:0.329979 12:0.603634 13:-0.368195 14:0.071696 15:0.921960 1
1.000000  1:-1.765220 2:0.434114 3:0.397484 4:-2.180240 5:-0.482221 6:0.579913 7:0.007180 8:-0.396361 9:-2.186180 10:0.360424 11:-0.479673 12:-0.808165 13:0.214944 14:0.760898 15:-0.334305 16
-1.000000  1:1.483440 2:0.567198 3:0.573307 4:-0.202935 5:1.108220 6:1.627170 7:1.202440 8:-0.093951 9:0.779752 10:0.540800 11:0.075543 12:-0.301512 13:-0.507922 14:-0.142293 15:1.049840 16:-
-1.000000  1:-0.953049 2:1.107940 3:1.328480 4:-1.644580 5:-0.074236 6:-0.776688 7:-1.405830 8:-0.338663 9:-0.583577 10:1.370500 11:0.195211 12:-0.159352 13:-1.425460 14:-0.232203 15:-0.2759
1.000000  1:0.214162 2:0.512283 3:0.318795 4:0.375563 5:-0.319845 6:-1.907150 7:-1.601020 8:0.697753 9:-0.008599 10:-0.606610 11:-0.098381 12:0.596746 13:-0.320061 14:1.313370 15:-0.072354 16
1.000000  1:0.426636 2:-1.864190 3:-1.847760 4:-0.258318 5:-1.380020 6:-0.065725 7:-0.268582 8:-0.769468 9:-2.508850 10:-1.138450 11:-2.049650 12:-1.219540 13:0.392029 14:1.048680 15:-0.6156
-1.000000  1:0.781703 2:-2.166330 3:-1.935410 4:-0.591776 5:-0.140932 6:-1.005580 7:-0.891690 8:-0.254486 9:0.518198 10:-1.455980 11:-0.476479 12:-0.536260 13:-0.144533 14:0.089240 15:0.15265
-1.000000  1:-0.333621 2:-0.331659 3:-0.039726 4:-0.684055 5:-1.182430 6:1.274620 7:0.962359 8:-1.101840 9:-0.192507 10:-0.438753 11:-0.670828 12:-1.316550 13:1.192580 14:0.026216 15:-0.7331
```

Normal text file | length : 1,727,461 lines : 63 | Ln : 1 Col : 1 Pos : 1 | Unix (LF) | UTF-8 | INS

پیش پردازش

اصلاح عددها و تبدیل cancer به فاکتور

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

پیش پردازش

اصلاح عددها و تبدیل cancer به فاکتور

Project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

main.R | note 1.Rmd | Paper I.R | server.R | Paper II.R | colon | ui.R | Paper V.R

Filter | Cols: « < 1 - 50 > »

	cancer	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10	Gene11	Gen
1	-1	2.080750	1.099070	0.927763	1.029080	-0.130763	1.265460	-0.436286	0.728881	2.107980	1.359870	0.265471	
2	1	1.109460	0.786453	0.445560	-0.146323	-0.996316	0.555759	0.290734	-0.145259	1.132660	0.559093	-1.469130	
3	-1	-0.676530	1.693100	1.559250	1.559980	-0.982179	-1.358510	-1.313990	-0.455067	0.295214	0.290694	0.415632	
4	1	0.534396	1.677540	1.489030	0.778605	-0.183776	-1.116850	-1.487560	-0.579511	0.292683	1.345480	-0.687898	
5	-1	-1.018900	0.511080	0.755641	1.013820	0.529899	0.160440	-0.087055	1.295290	0.458736	0.714082	0.727290	
6	1	-1.185370	-0.514473	-0.566634	1.224720	0.619244	-0.684713	-0.798129	1.368770	-0.697007	-1.006190	0.808748	
7	-1	1.779050	0.423947	0.820696	2.525690	0.666921	0.661346	0.425365	0.165247	1.967910	1.140080	-0.779230	
8	1	-0.889638	-0.315453	-0.073131	1.157500	-0.311039	-0.364472	-1.621640	1.193000	0.689805	0.203786	1.907460	
9	-1	-0.659694	-0.184388	-0.540022	1.122420	0.562609	-2.988310	-2.349810	-1.325010	-0.017002	0.109081	1.907460	
10	1	-1.225800	-0.212615	-0.588923	1.335410	-0.356505	0.354394	0.699607	0.190782	-0.139117	-0.551846	0.292320	
11	-1	-0.377282	-2.620490	-2.763910	0.612038	-0.155718	-1.456070	0.683292	-1.031570	0.718861	-2.313070	-0.092949	
12	1	1.639210	0.663551	0.324924	1.456460	-0.170587	-0.443486	-0.832590	-0.928334	1.320070	0.591865	0.158091	

Showing 1 to 13 of 62 entries, 2001 total columns

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

Paper I
- Breast and Colon Cancer Classification from Gene Expression Profiles Using Data Mining Techniques
- GW, IG, SVM

Paper II
- Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques
- RF, MDA, MDG, CV, Ensemble

Paper III
- Examining Applying High Performance Genetic Data Feature Selection and Classification Algorithms for Colon Cancer Diagnosis
- PSO, SVM

Paper IV
- Naïve Bayes Classifier Models for Predicting the
- Colon Cancer
- Naïve Bayes

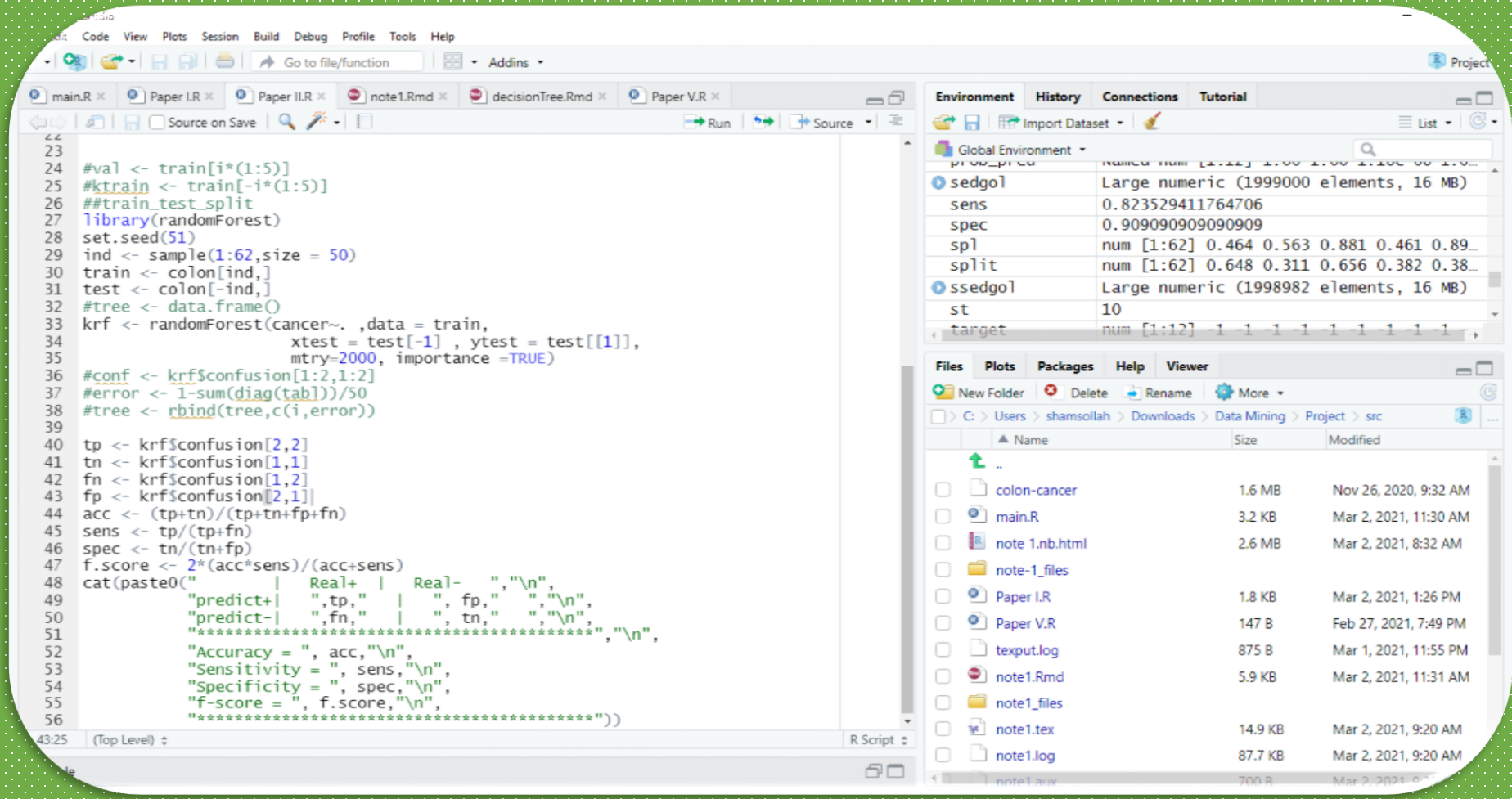
Paper V
- Statistical Characterization and Classification of Colon Microarray Gene Expression Data using Multiple Machine Learning Paradigms
- Jack of all trades!

Packages
- Advanced Bioinformatics
- Special Packages in R

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

- Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques
- RF, MDA, MDG, CV, Ensemble



مدلسازی

- Naïve Bayes Classifier Models for Predicting the Colon Cancer
- Naïve Bayes

فاعده بیز

$$\Pr(C|X) = \frac{\Pr(X|C)\Pr(C)}{\Pr(X)}$$

که با جایگذاری ویژگی ها به جای X به فرمول زیر می رسیم:

$$\Pr(C|x_1, x_2, \ldots, x_{2000}) = \frac{\Pr(x_1, x_2, \ldots, x_{2000}|C)\Pr(C)}{\Pr(x_1, x_2, \ldots, x_{2000})}$$

با فرض استقلال هر دو ویژگی از هم داریم:

$$\Pr(C|x_1, x_2, \ldots, x_{2000}) = \Pr(C)\Pi_{i=1}^{2000}\Pr(x_i|C)$$

تا اینجا همه چیز مطابق چیزی بود که در درس داشتیم اما قسمت متفاوت رفتار با داده های پیوسته است که پیشنهاد مقاله این است که برای آنها احتمال شرطی را به صورت زیر حساب می کنیم:

$$\Pr(X_i = x_i|C = c_j) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

حال می توانیم دست به کد شویم:

مدلسازی

- Naïve Bayes Classifier Models for Predicting the Colon Cancer
- Naïve Bayes

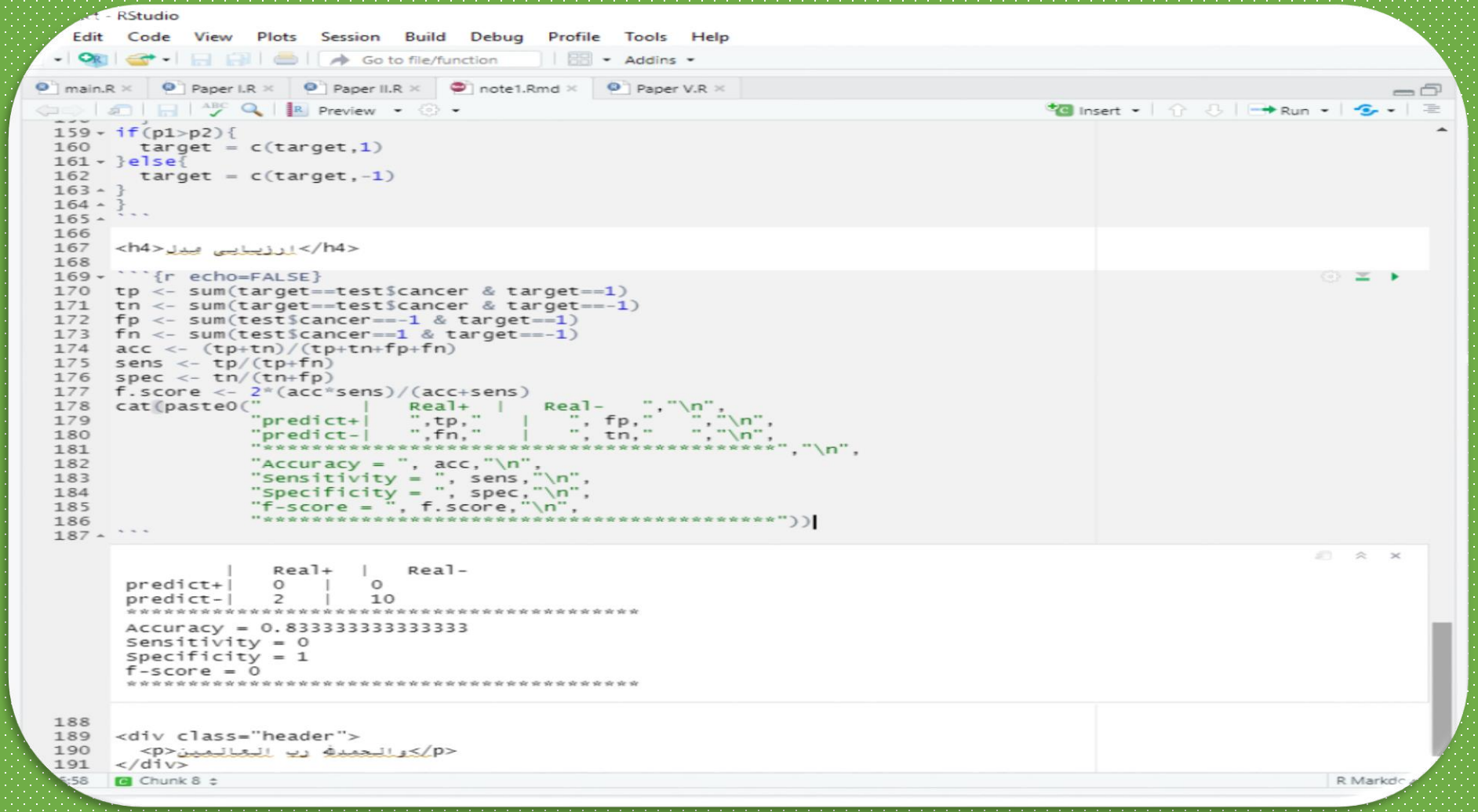
```
cancer probability
cprob <- sum(train$cancer==1)/dim(train)[1]
hprob <- 1-cprob
ctrain <- subset(train,train$cancer==1)
htrain <- subset(train,train$cancer==-1)
hmean <- sapply(htrain[,2:2001], mean)
hsd <- sapply(htrain[,2:2001], sd)
cmean <- sapply(ctrain[,2:2001], mean)
csd <- sapply(ctrain[,2:2001], sd)
target = c()
for (i in 1:nrow(test)) {
p1 = cprob
p2 = hprob
  for (j in 1:2000) {
    p1 = p1*dnorm(test[[i,j+1]],cmean[j],csd[j],log = F)
    p2 = p2*dnorm(test[[i,j+1]],hmean[j],hsd[j],log = F)
  }
if(p1>p2){
  target = c(target,1)
}else{
  target = c(target,-1)
}
```


- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

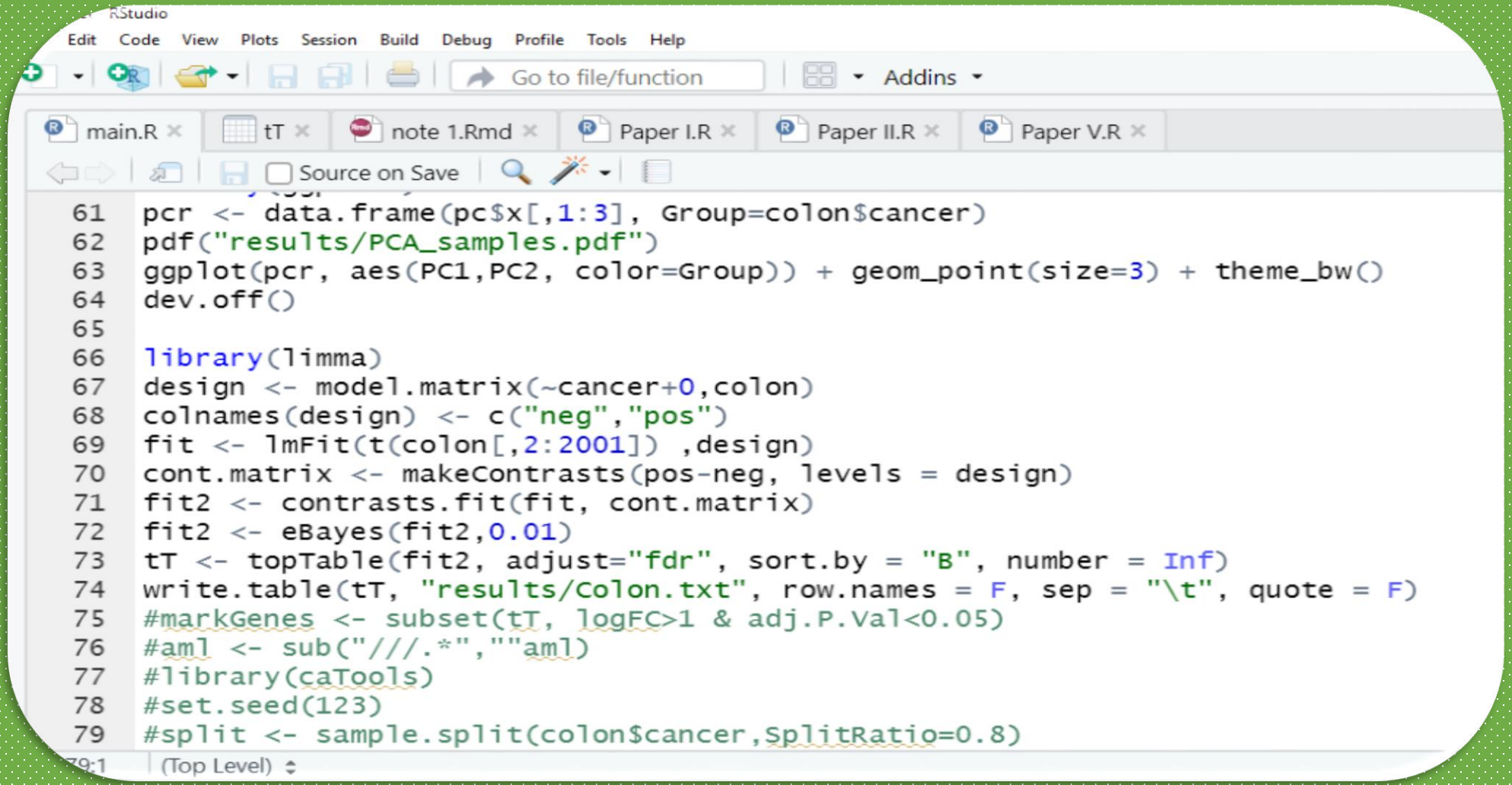
- Naïve Bayes Classifier Models for Predicting the Colon Cancer
- Naïve Bayes



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

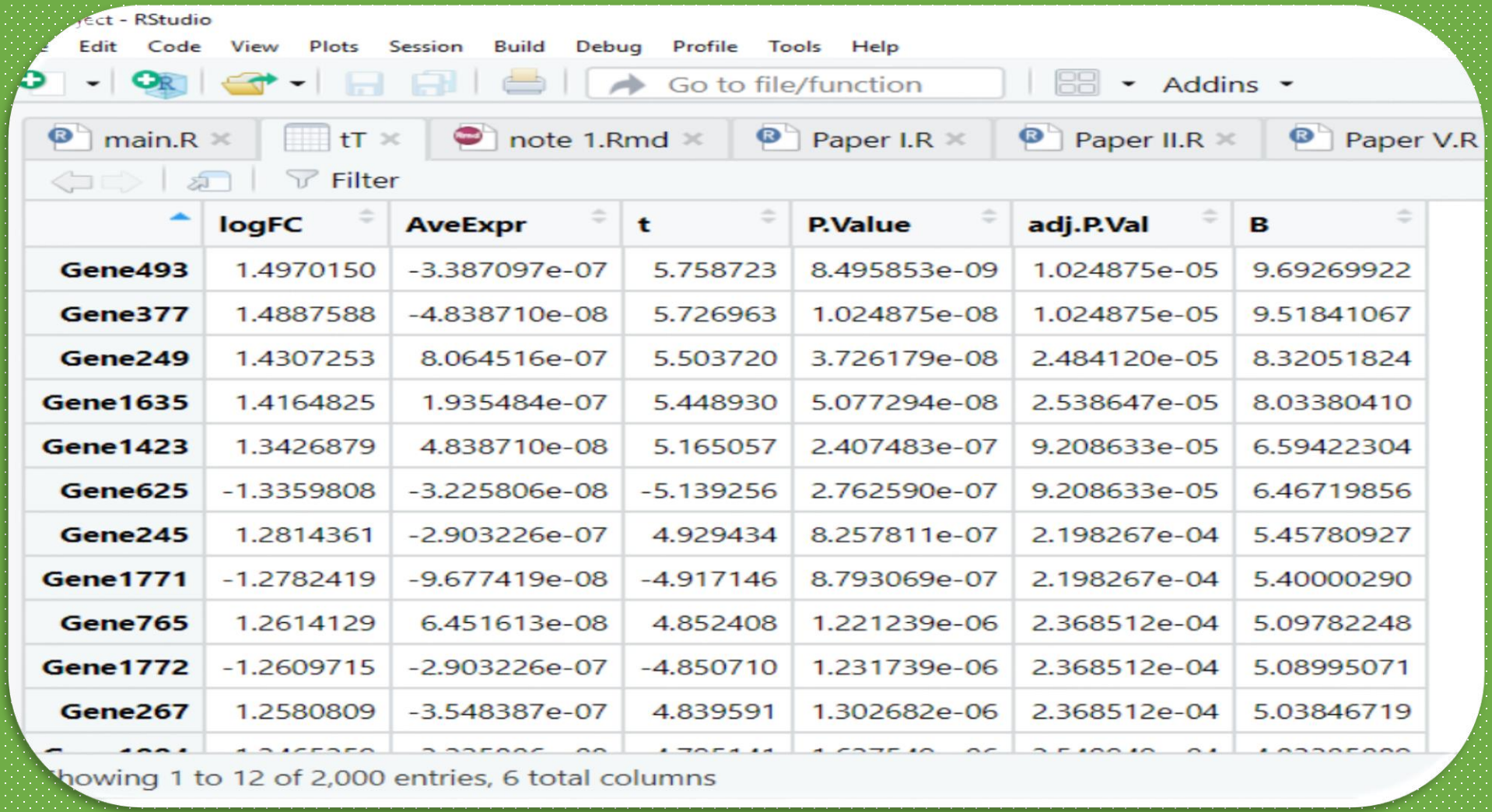
- Advanced Bioinformatics
- Special Packages in R



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

- Advanced Bioinformatics
- Special Packages in R

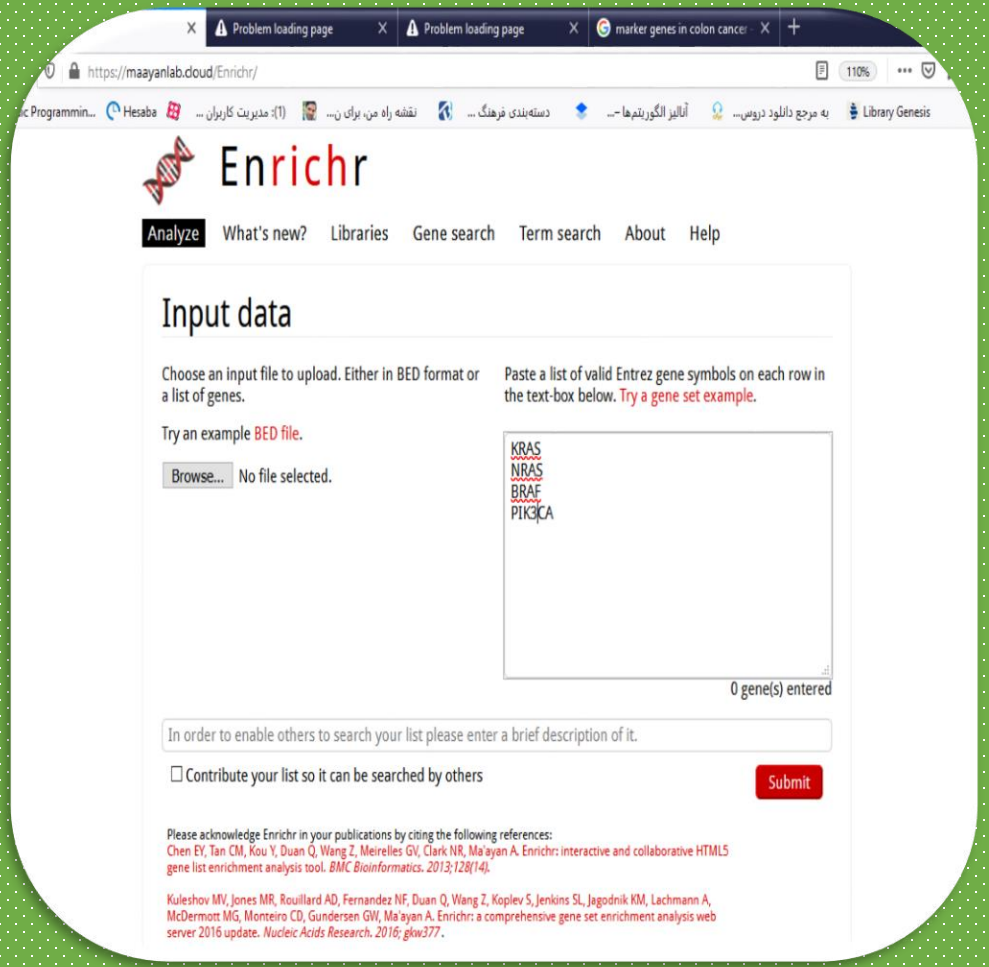


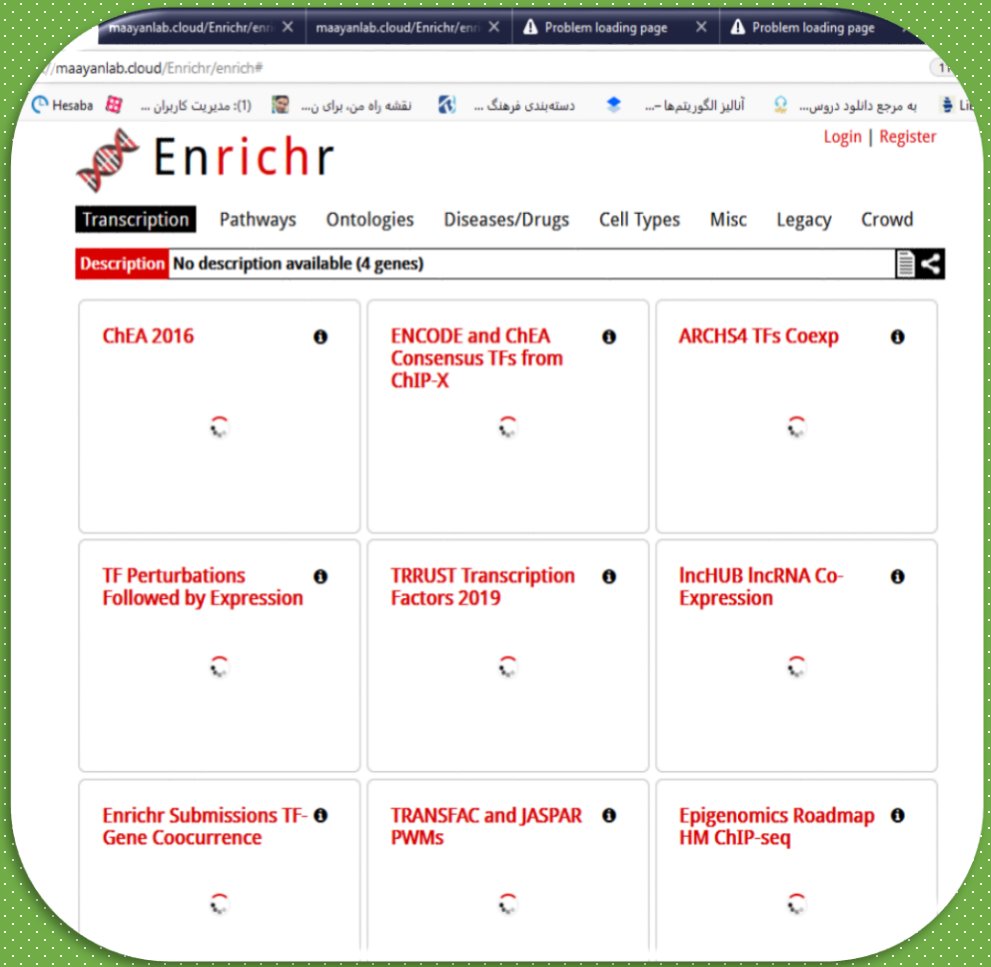
	logFC	AveExpr	t	P.Value	adj.P.Val	B
Gene493	1.4970150	-3.387097e-07	5.758723	8.495853e-09	1.024875e-05	9.69269922
Gene377	1.4887588	-4.838710e-08	5.726963	1.024875e-08	1.024875e-05	9.51841067
Gene249	1.4307253	8.064516e-07	5.503720	3.726179e-08	2.484120e-05	8.32051824
Gene1635	1.4164825	1.935484e-07	5.448930	5.077294e-08	2.538647e-05	8.03380410
Gene1423	1.3426879	4.838710e-08	5.165057	2.407483e-07	9.208633e-05	6.59422304
Gene625	-1.3359808	-3.225806e-08	-5.139256	2.762590e-07	9.208633e-05	6.46719856
Gene245	1.2814361	-2.903226e-07	4.929434	8.257811e-07	2.198267e-04	5.45780927
Gene1771	-1.2782419	-9.677419e-08	-4.917146	8.793069e-07	2.198267e-04	5.40000290
Gene765	1.2614129	6.451613e-08	4.852408	1.221239e-06	2.368512e-04	5.09782248
Gene1772	-1.2609715	-2.903226e-07	-4.850710	1.231739e-06	2.368512e-04	5.08995071
Gene267	1.2580809	-3.548387e-07	4.839591	1.302682e-06	2.368512e-04	5.03846719

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

- Advanced Bioinformatics
- Special Packages in R

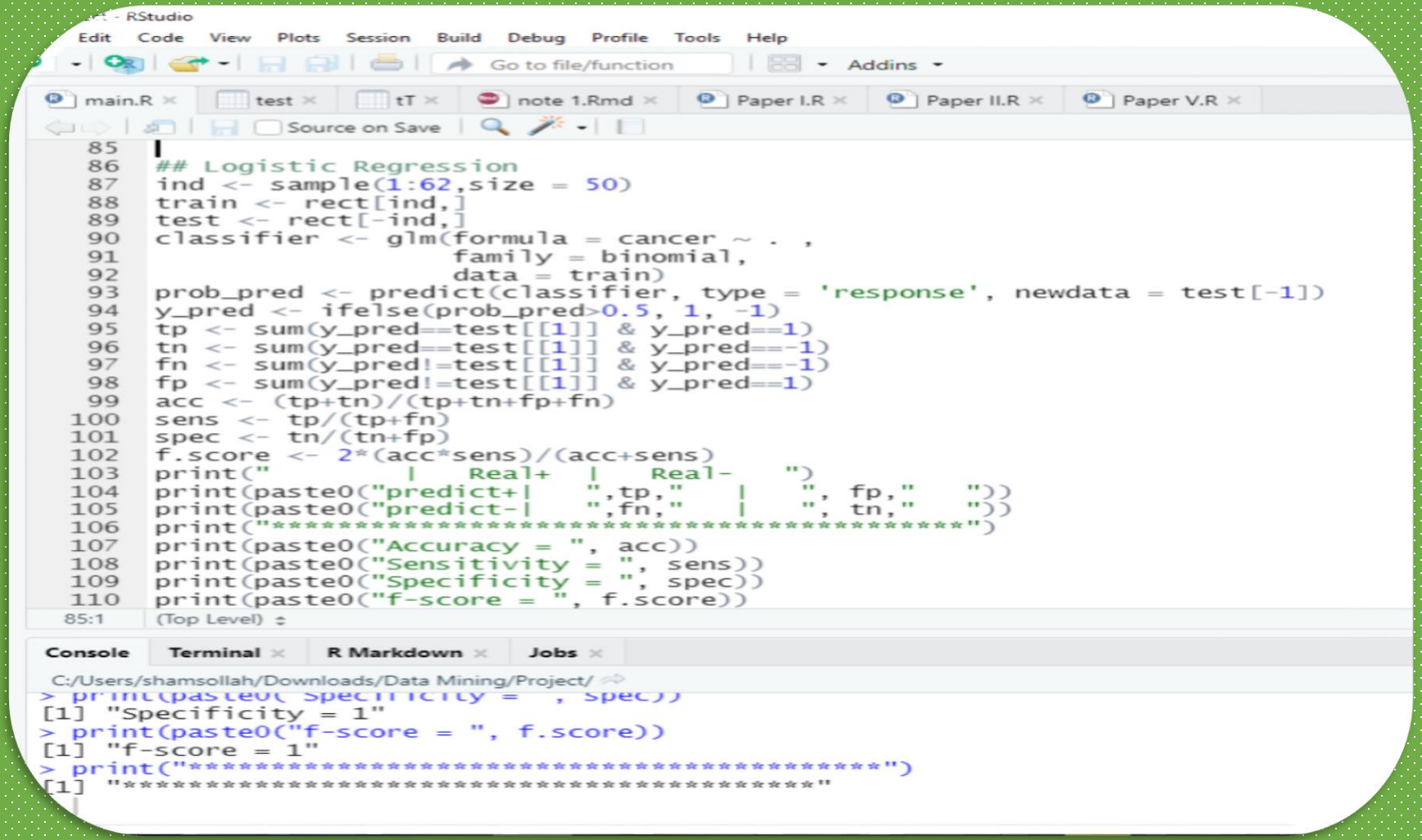




- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مدلسازی

- Advanced Bioinformatics
- Special Packages in R

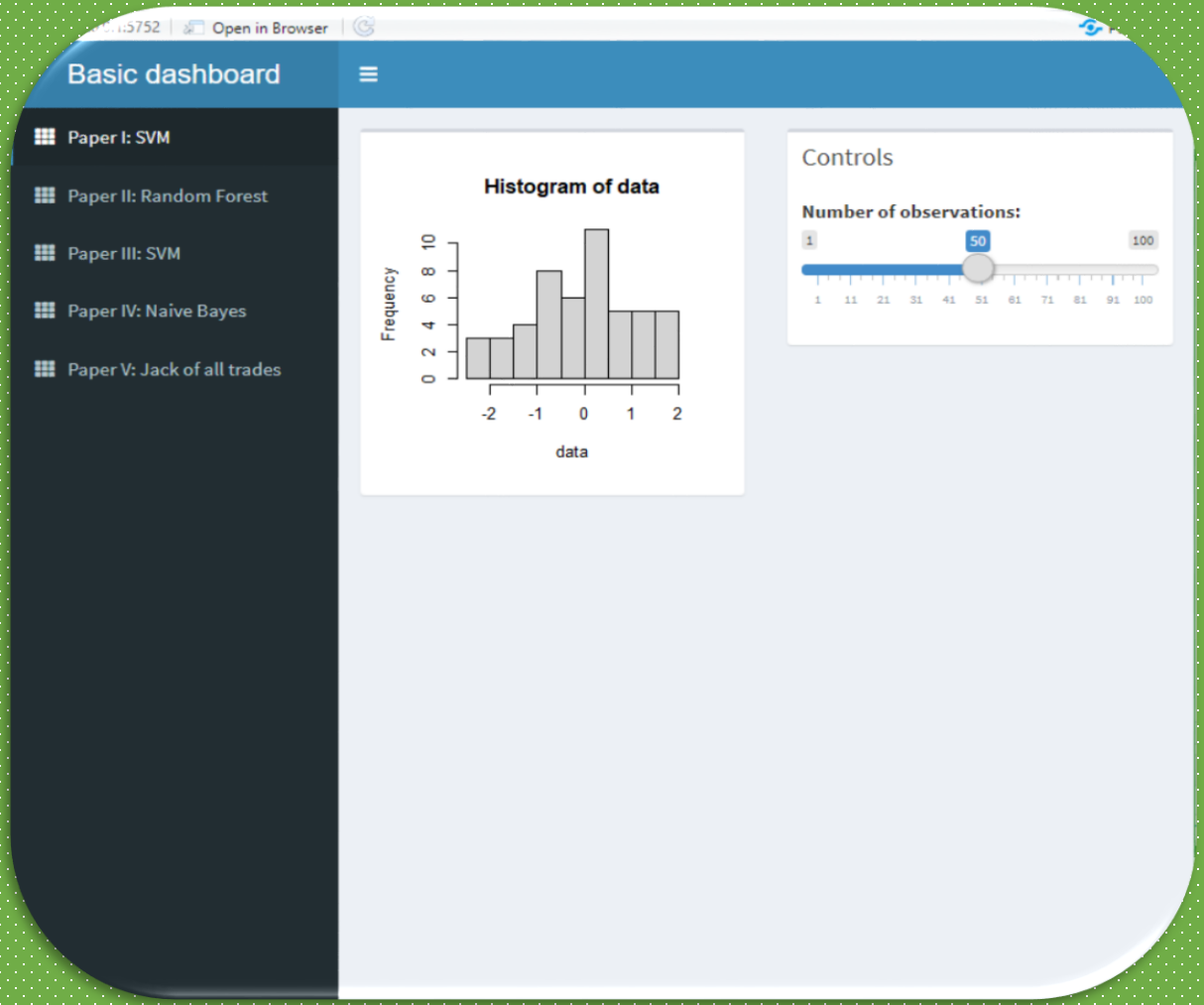


ارزیابی

- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

Method	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	F-Score
RF	14	3	30	3	88	82.35	90.90	85
NB	0	0	10	2	83.33	0	100	0
SVM	8	0	4	0	100	100	100	100
Spec	5	0	7	0	100	100	100	100

ارائه



- ابزار کار
- خواندن داده
- پیش پردازش
- مدلسازی
- ارزیابی
- ارائه

مراجع

- http://ce.sharif.edu/courses/97-98/2/ce550-1/
- maktabkhooneh.org
- Foundations and Applications of Statistics An Introduction Using R [1,2,3]
- R for data science

با سپاس از توجه شما