Toggle code

Jiwei and Xueer's Insight: Beyond the leaks

This study is done by Jiwei Liu and Xueer Chen, part of the team better luck in 2016, 6th place of the Genentech Cervical Cancer Screening contest. We first study the impact of leaks, absence of a claim line in either diagnosis_head or procedure_head in the datasets, and show that their predicting power is dominant. Therefore in the following analysis we remove all the records with leak features to better identify other important features. We modified Follow the Regularized Leader (FTRL) algorithm to extract important features and feature interactions in an online fashion. Our approach can find both leak and good features automatically without any prior domain knowledge.

In summary, we make the following insights:

1. The leak is frequent and strong.

31.8% Patients have leak features and 97.8% of them are screeners. So we remove the records with leak to do the following analysis.

2. Most useful features are code features: diagnosis code, procedure code, surgical code and their interactions.

Diagnosis code and procedure code pairs linked by the same claim ID can be a strong indicator of screeners. Specifically, the algorithm found 109 code pairs that correspond to more than 90% screener rate.

A single diagnosis code, procedure code or surgical code is also a good indicator of screeners but much more frequent in data sets. The algorithm found 27 such single codes that correspond to more than 80% screener rate.

A single procedure code or a code pair can be a good indicator of non-screeners. The algorithm found 4 codes and 5 pairs which correspond to less than 30% screener rate and more than 0.01% frequency.

3. There are some "hidden" examinations that might indicate the necessity of cervical cancer examination.

"ASSAY OF THYROID STIMULATING HORMONE TSH" is kind of surprising as this test is usually for evaluating thyroid function rather than cervical cancer relevant test; our model recognizes that people have this test are more likely to be screeners; out of curiosity, we found a research on "Expression of Thyroid Stimulating Hormone in Cervical Epithelial Cells" indicating that these two are related, they hypothesized that TSH may play a role in the mechanism of action of human papilloma virus infection in cervical epithelia, and subsequent tumorigenesis.

4. People with acute diseases are less likely to have screeners.

We can explain this by a "mutual exclusive" concept, i.e., people already occupied with acute/chronic diseases are not likely to have the necessity or efforts to care about other severe diseases (cervical cancer), either the possibility they have cervical cancer given they already have other acute diseases are lower, or they don't have the efforts to realize they need to do regular screening.

5. Special populations might need help on screening.

We find people with mental diseases or gender identity disorder are less likely to be screeners; for people have mental diseases, they may have difficulties recognizing the necessity of having examination themselves, thus people who take care of them such as families or healthcare professionals should help them; for people with gender identity disorder, they also need appropriate education on proper examination for them.

6. Some practitioners' job directly relates to screening or cervical cancer.

For example, practitioner "16444459", "12709441", and "14015592" are doing DIAGNOSTIC RADIOLOGY and OBSTETRICS AND GYNECOLOGY and they correspond to +90% screener rate!

7. Nurse's experience matters

For example, NMW, NURSE MIDWIFE, advanced practice registered nurses, correspond to 81% screener rate while SNP, SCHOOL NURSE PRACTITIONER, correspond to only 25% screener rate

8. We show that the features extracted by the algorithm achieves good predictive power accross different age group, ethinicity, education level and income.

Preprocess the data, using pypy.

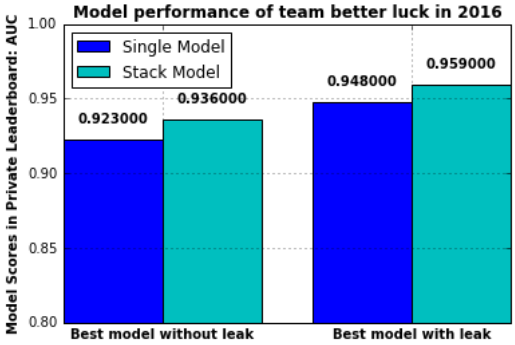
0

Insight 1: Leak is dominant

Model performance

Unfortunately there is leak in this data set. So how bad is it?

In our case, before finding the leak, our best single model and best stacking model achieve 0.936 auc score in both public and private leaderboard. After finding the leak, our best stacking model gets 0.959 auc. At that time we didn't realize what we found is a leak (we thought it is a golden feature) and we failed to fully exploit it like other top teams.



Conclusion: the leak will boost the model auc score by at least 0.02.

A closer look at leak

We extract 400 most important features using ftrl from diagnosis_head.csv and procedure_head.csv, which only include occurrences of diagnosis codes, procedure codes and their interactions. With these 400 features, a ftrl model can get 0.908 auc in less than 10 minutes. Features are extracted based on their weight in the ftrl model: the greater absolute value of the weight, the more important the feature.

The features can be divided into three categories:

1. the occurrence of a single diagnosis/procedure code
2. the occurrence of a pair of a diagnosis code and a procedure code connected by the same claim id.
3. the occurrence of a pair of a diagnosis code and a missing procedure code connected by the same claim id.

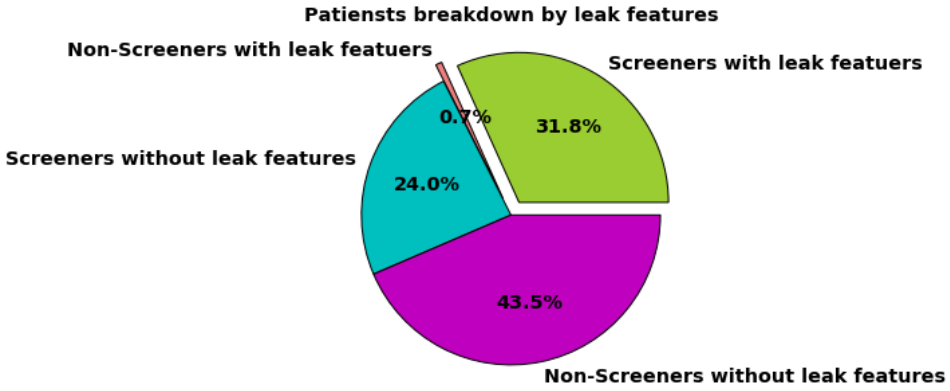
As we know, type 3) the missing code or entire procedure entries are actually leaks.

```
-c:1: FutureWarning: sort(columns=....) is deprecated, use sort_values(by=.....)
```

	index	feature	weight	screen_rate	frequency	description	is_leak	is_positive
162	167930709	V72.31	0.557454	0.825552	0.492814	ROUTINE GYNECOLOGICAL EXAMINATION	0	True
253	253625570	88305	0.117017	0.643462	0.373726	p:LEVEL IV SURG PATHOLOGY GROSS&MICROSCOPIC EXAM	0	True
87	27787726	V70.0	0.105387	0.695280	0.302872	ROUTINE GENERAL MEDICAL EXAMINATION AT A HEALTH CARE FACILITY	0	True
164	141421400	99396	0.111948	0.728219	0.276361	p:PERIODIC PREVENTIVE MED EST PATIENT 40-64YRS	0	True
86	105611720	V72.31_	2.977993	0.998394	0.233918	ROUTINE GYNECOLOGICAL EXAMINATION	1	True
267	21665063	V72.31_99396	-0.407650	0.775715	0.169847	ROUTINE GYNECOLOGICAL EXAMINATION |||| PERIODIC PREVENTIVE MED EST PATIENT 40-64YRS	0	False
266	21665062	V72.31_99395	-0.333932	0.839669	0.100202	ROUTINE GYNECOLOGICAL EXAMINATION |||| PERIODIC PREVENTIVE MED EST PATIENT 18-39 YRS	0	False
271	105844015	87491	0.658191	0.905852	0.098132	p:IADNA CHLAMYDIA TRACHOMATIS AMPLIFIED PROBE TQ	0	True
327	187770410	87591	0.640004	0.905743	0.095544	p:IADNA NEISSERIA GONORRHOEAE AMPLIFIED PROBE TQ	0	True
181	210008996	87081	0.113113	0.779744	0.073474	p:CUL PRSMPTV PTHGNC ORGANISM SCRN W/COLONY ESTIMJ	0	True

feamap8.csv is the feature selection result from ftrl which include important codes and codes interactions. All original occurrences of codes and their interactions are mapped/hashed to 2^28 feature space.

1. The column 'index' indicates the feature's index in the 2^28 feature space.
2. The column 'feature' indicates the content of the feature, from which we can tell the 3 types. For example, V77.71 is the 1st type or occurrence of a single diagnosis code. '*V76.41G0328' is the 2nd type or a pair of codes. V76.41 is a diagnosis code and G0328 is a procedure code and they are connected by the same claim_id. 'V88.01*' is the 3rd type or a leak feature. V88.01 is a diagnosis code but the corresponding procedure code with the same claim_id is missing.
3. The column 'weight' stands for the feature weight in ftrl model. A positive/negative weight indicates a screener/non-screener accordingly.
4. The column 'screen_rate' indicates the percentage of screeners in patients whose records include this feature at least once.
5. The column 'frequency' indicates the percentage of patients whose records include this feature at least once out of all the patients
6. The column 'description' includes the descriptions of the codes. For code pairs descriptions are in the format of "diagnosis code description |||| procedure code description"
7. The column 'is_leak' indicates whether the feature is a leak feature. (diagnosis_code + missing procedure code of a claim)
8. The column 'is_positive' indicates if a feature has a positive weight




```
number of leak features 79 , number of non-leak features 325
```

This figure shows that 31.8% patients who have leak features in their records are also screeners and only 0.7% patients who have leak features are not screeners. In other words, 97.8% patients who have leak features are screeners. Since we didn't consider another leak case in this part, that a claim is present in procedure but not diagnosis, the actual patients with leak features could be more than the figure showed.

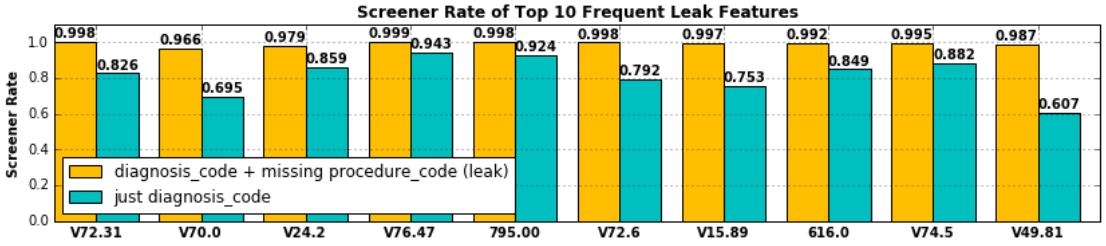
Out of 404 important features 19.6% (79) are leak features, out of which 97.5% leak features indicate screeners.

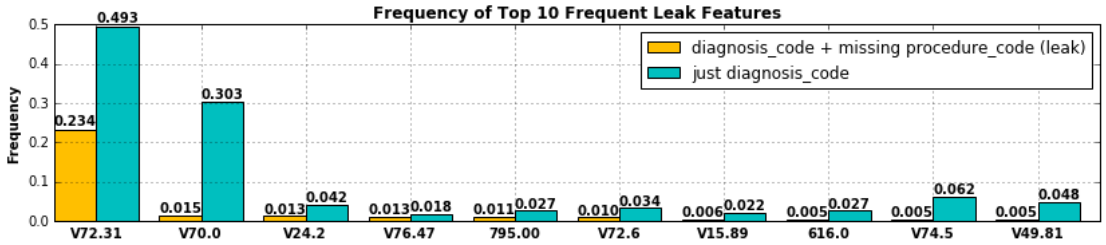
Next, let's find out what code pairs indicate strong leak.

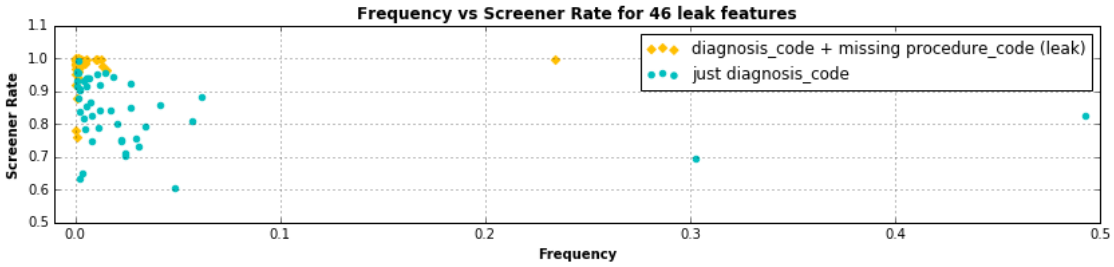
```
(79, 8) (46, 11)
```

```
-c:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```








```
mean screener rate for leak features 0.975095478261 summed frequency of leakfeatures 0.356516
```

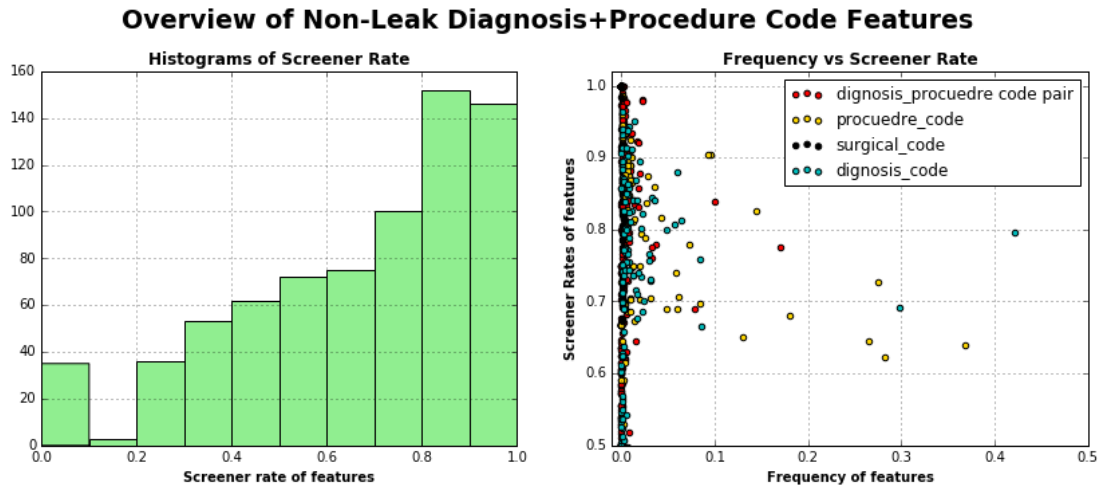
These figures show that the leak features strongly indicate screeners, and their predicting power are much stronger than the occurrence of the same diagnosis code without considering the missing pair (labelled as "just diagnosis code" in the figure). Please note that the green bar includes the yellow bar for each code. For example, 23% patients have the leak feature with diagnosis code V72.31, and 99.8% of them are screeners! For all 76 leak features, the average screener rate is 97.5% while non-leak features has the average screener rate 84%. Except for diagnosis code V72.31, other leak features are not frequent, smaller than 3%.

Insight from the "Frequency vs Screener Rate For 46 leak features": taken the three points with a frequency higher than 0.2, they are V72.31_ (routine gynecological examination combined with a missing feature), V72.31 (routine gynecological examination) and V70.0(routine gynecological examination). V72.31 itself has a high screener_rate (0.826), however, by combining with an unknown feature, the screener_rate is even higher (0.998), we can assume that the unknown feature can be some examination or procedure that are fairly relevant to cervical cancer (it can even be a HPV test). It turns out people who take routine gynecological examination are more likely to take cervical-related procedure than who take routine general medical examination (V70.0).

Insight 2: Most useful features are code features: diagnosis code, procedure code, surgical code and code pairs

```
(734, 8) (734, 8)
```

```
-c:2: FutureWarning: the take_last=True keyword is deprecated, use keep='last' instead
-c:2: FutureWarning: the 'cols' keyword is deprecated, use 'subset' instead
-c:3: FutureWarning: sort(columns=....) is deprecated, use sort_values(by=.....)
```



First, let's take a close look at the procedure codes with frequency greater than 0.1

	index	feature	weight	screen_rate	frequency	description	is_leak	is_positive	feature_type
269	141421397	99395	0.182217	0.826148	0.145080	p:PERIODIC PREVENTIVE MED EST PATIENT 18-39 YRS	0	True	procedure_code
270	141421400	99396	0.159637	0.727765	0.275899	p:PERIODIC PREVENTIVE MED EST PATIENT 40-64YRS	0	True	procedure_code
300	210008995	87086	0.094313	0.680451	0.180286	p:CULTURE BACTERIAL QUANTTATIVE COLONY COUNT URINE	0	True	procedure_code
441	192859443	82306	0.086656	0.650494	0.129413	p:25 HYDROXY INCLUDES FRACTIONS IF PERFORMED	0	True	procedure_code
354	139713571	84443	0.088933	0.645343	0.265162	p:ASSAY OF THYROID STIMULATING HORMONE TSH	0	True	procedure_code
411	253625570	88305	0.136199	0.638884	0.368906	p:LEVEL IV SURG PATHOLOGY GROSS&MICROSCOPIC EXAM	0	True	procedure_code
326	9827293	80061	0.097111	0.623372	0.281992	p:LIPID PANEL	0	True	procedure_code

Insight 3: There are some "hidden" examinations that might indicate the necessity of cervical cancer examination.

Taken from table above, we found the following features interesting:

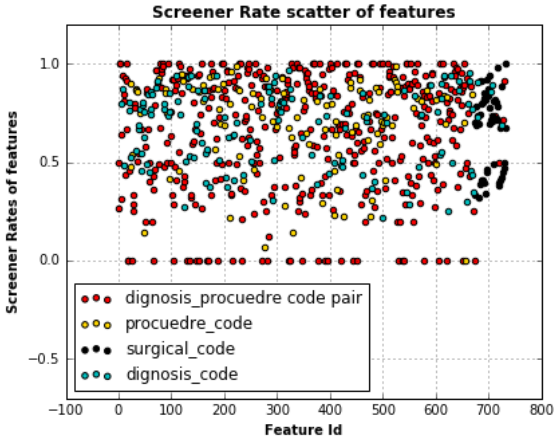
1. PERIODIC PREVENTIVE MED EST PATIENT
2. ASSAY OF THYROID STIMULATING HORMONE TSH
3. LEVEL IV SURG PATHOLOGY GROSS&MICROSCOPIC EXAM
4. LIPID PANEL

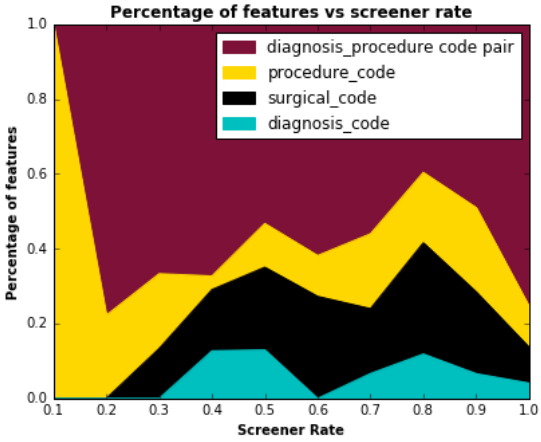
The insights are:

1. Women who have "PERIODIC PREVENTIVE MED EST PATIENT" are more likely to be screeners
2. "ASSAY OF THYROID STIMULATING HORMONE TSH" is kind of surprising as this test is usually for evaluating thyroid function rather than cervical cancer relevant test; our model recognizes that people have this test are more likely to be screeners; out of curiosity, we found a research on "Expression of Thyroid Stimulating Hormone in Cervical Epithelial Cells" indicating that these two are related, they hypothesized that TSH may play a role in the mechanism of action of human papilloma virus infection in cervical epithelia, and subsequent tumorigenesis. [ref: http://oaktrust.library.tamu.edu/handle/1969.1/154992]
3. "LEVEL IV SURG PATHOLOGY GROSS&MICROSCOPIC EXAM" may indicate that patients have gone through (cervical-related) sugeries [ref: http://a1.mayomedicallaboratories.com/webjc/attachments/53/05c49d7-surgical-pathology.pdf]
4. "LIPID PANEL" is a blood test that might be prescribed to patients to check their cervical carcinoma status

```
All:734, diagnosis_procedure code pair:438, diagnosis_code:135, procedure_code:111, surgical_code:50
```

The histogram plot shows the distribution of screener rates of all features. More than 50% of the features correspond to screener rate greater than 0.8. There are also 10% features that strongly indicate non-screeners





The insight:

scatter plot gives a better view of the distribution of these features. It shows that code pairs tend to have extremely high a nd low screener rate. Part of this is due to the fact that such pairs are very infrequent in the datasets. Most diagnosis_code correspond to screener rate greater than 0.5. Most features that correspond to screener rate less than 0.5 are procedure codes and code pairs. Some surgical codes also strongly indicate screeners.

The area plot on the right shows the percentage of each type of features at different screener rate zone. It clearly shows tha t in screener rate zone [0,0.2] procedure codes and code pairs are the majority and they strongly indicate non-screeners while all of 4 types have codes that strongly indicate screeners, out of which code pairs is the majority.

Top 10 Strong and Frequent Indicators of Screeners

```
-c:2: FutureWarning: sort(columns=....) is deprecated, use sort_values(by=.....)
```

	feature	screen_rate	description	frequency	feature_type
244	V72.31_87491	0.980587	ROUTINE GYNECOLOGICAL EXAMINATION |||| IADNA CHLAMYDIA TRACHOMATIS AMPLIFIED PROBE TQ	0.022601	code_pair
341	V72.31_87591	0.978956	ROUTINE GYNECOLOGICAL EXAMINATION |||| IADNA NEISSERIA GONORRHOEAE AMPLIFIED PROBE TQ	0.022122	code_pair
663	622.11	0.950659	MILD DYSPLASIA OF CERVIX	0.013373	diagnosis_code
537	V22.1_87491	0.934735	SUPERVISION OF OTHER NORMAL PREGNANCY |||| IADNA CHLAMYDIA TRACHOMATIS AMPLIFIED PROBE TQ	0.010309	code_pair
75	57454	0.925105	p:COLPOSCOPY CERVIX BX CERVIX & ENDOCRV CURRETAGE	0.010091	procedure_code
356	V74.5_87591	0.923598	SCREENING EXAMINATION FOR VENEREAL DISEASE |||| IADNA NEISSERIA GONORRHOEAE AMPLIFIED PROBE TQ	0.017070	code_pair
433	V74.5_87491	0.921770	SCREENING EXAMINATION FOR VENEREAL DISEASE |||| IADNA CHLAMYDIA TRACHOMATIS AMPLIFIED PROBE TQ	0.017488	code_pair
664	622.10	0.912395	DYSPLASIA OF CERVIX, UNSPECIFIED	0.010766	diagnosis_code
439	87491	0.903392	p:IADNA CHLAMYDIA TRACHOMATIS AMPLIFIED PROBE TQ	0.095329	procedure_code
525	87591	0.903227	p:IADNA NEISSERIA GONORRHOEAE AMPLIFIED PROBE TQ	0.092757	procedure_code

This table shows the top 10 strong and frequent indicators: these indicators agree with common sense as they are all related to cervical cancer. Most of them are directly related to cervix or screening.

Top 10 Strong and Frequent Indicators of Non-Screeners

	feature	screen_rate	description	frequency	feature_type
211	G0164	0.221326	p:SKILLED SERVICES OF A LICENSED NURSE (LPN OR RN)	0.000964	procedure_code
426	A0800	0.222222	p:AMB TRANSPORT PROV BETWN HR 7PM&7AM	0.000132	procedure_code
405	585.6_83540	0.256140	END STAGE RENAL DISEASE :::: ASSAY OF IRON	0.001477	code_pair
0	585.6_84460	0.264666	END STAGE RENAL DISEASE :::: TRANSFERASE ALANINE AMINO ALT SGPT	0.000633	code_pair
597	250.42_82947	0.270440	DIABETES MELLITUS WITH RENAL MANIFESTATIONS, TYPE II OR UNSPECIFIED TYPE, UNCONTROLLED :::: GLUCOSE QUANTITATIVE BLOOD XCPT REAGENT STRIP	0.000137	code_pair
313	B4155	0.280000	p:ENTERAL FORMULA, NUTRITIONALLY INCOMPLETE/MODULA	0.000130	procedure_code
27	V58.69_97110	0.284960	LONG-TERM (CURRENT) USE OF OTHER MEDICATIONS :::: THERAPEUTIC PX 1/> AREAS EACH 15 MIN EXERCISES	0.000655	code_pair
386	90779	0.295559	p:UNLISTED THERAPEUTIC, PROPHYLACTIC OR DIAGNOSTIC INTRAVENOUS OR INTRA-ARTERIAL INJECTION OR INFUSION	0.000564	procedure_code
342	518.81_99308	0.299270	ACUTE RESPIRATORY FAILURE :::: SBSQ NURSING FACIL CARE/DAY MINOR COMPLJ 15 MIN	0.000118	code_pair
291	797	0.304348	SENILITY WITHOUT MENTION OF PSYCHOSIS	0.000179	diagnosis_code

Insight 4: People with acute diseases are less likely to have screeners.

The table above lists top 10 strong and frequent indicators of non-screeners, there are some interesting findings:

```
1. people with acute or chronic diseases (not related to cervical cancer) are less likely to be screeners, we can explain this
by a "mutual exclusive" concept, i.e., people already occupied with acute/chronic diseases are not likely to have the necessit
y or efforts to care about other severe diseases (cervical cancer), either the possibility they have cervical cancer given the
y already have other acute diseases are lower, or they don't have the efforts to realize they need to do regular screening.
2. regarding the procedure code, the "AMB TRANSPORT PROV BETWN HR 7PM&7AM" may just indicate some emergency patients who have
accidents rather than diseases, which means they don't necessarily need a screen.
```

Top 10 Strong and Frequent procedure_code of Non-Screeners

	feature	screen_rate	description	frequency	feature_type
211	G0164	0.221326	p:SKILLED SERVICES OF A LICENSED NURSE (LPN OR RN)	0.000964	procedure_code
386	90779	0.295559	p:UNLISTED THERAPEUTIC, PROPHYLACTIC OR DIAGNOSTIC INTRAVENOUS OR INTRA-ARTERIAL INJECTION OR INFUSION	0.000564	procedure_code
426	A0800	0.222222	p:AMB TRANSPORT PROV BETWN HR 7PM&7AM	0.000132	procedure_code
313	B4155	0.280000	p:ENTERAL FORMULA, NUTRITIONALLY INCOMPLETE/MODULA	0.000130	procedure_code

Top 10 Strong and Frequent diagnosis_code of Non-Screeners

	index	feature	weight	screen_rate	frequency	description	is_leak	is_positive	feature_type	feature_id
549	142210700	302.85	-0.080127	0.2875	0.000069	GENDER IDENTITY DISORDER IN ADOLESCENTS OR ADULTS	0	False	diagnosis_code	549
500	131210734	302.50	-0.093708	0.2200	0.000043	TRANS-SEXUALISM WITH UNSPECIFIED SEXUAL HISTORY	0	False	diagnosis_code	500
123	99424651	331.7	-0.081206	0.2750	0.000035	CEREBRAL DEGENERATION IN DISEASES CLASSIFIED ELSEWHERE	0	False	diagnosis_code	123
623	32423764	201.43	-0.095377	0.2500	0.000007	HODGKIN'S DISEASE, LYMPHOCYTIC-HISTIOCYTIC PREDOMINANCE INVOLVING INTRA-ABDOMINAL LYMPH NODES	0	False	diagnosis_code	623

Insight 5: Special populations might need help on screening.

From the top strong and frequent diagnosis_code for non-screeners, we have some ineresting insights:

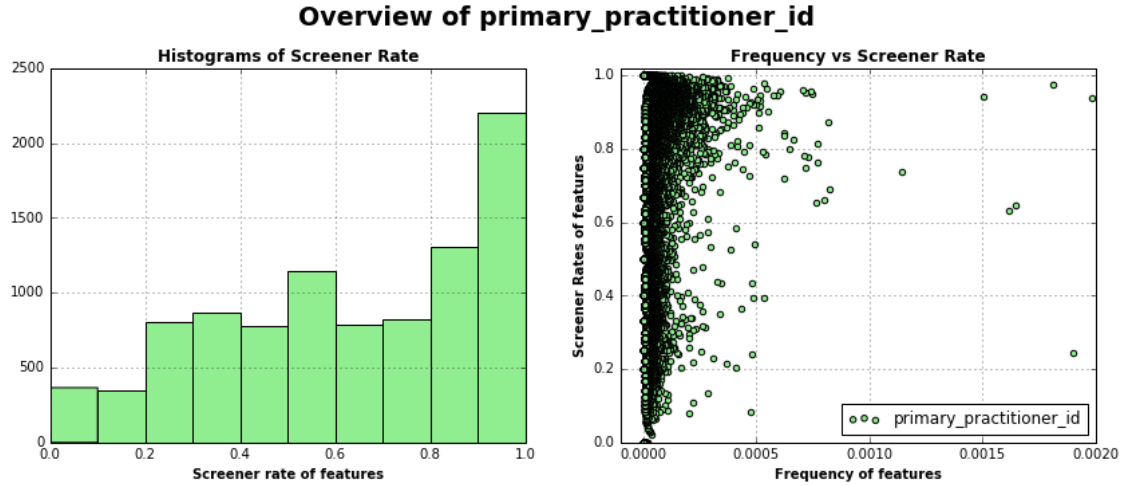
1. "GENDER IDENTITY DISORDER IN ADOLESCENTS OR ADULTS" and "TRANS-SEXUALISM WITH UNSPECIFIED SEXUAL HISTORY", these two indicators might represent a population that might need specific education, i.e., people with gender identity disorder or inclear recognition of their gender may have problems realizing the necessity and importance of having screenings; proper education should be transported to them regarding cervical cancer screening.

2. "CEREBRAL DEGENERATION IN DISEASES CLASSIFIED ELSEWHERE" may represent patients with mental diseases, which indicates they may not have enough sense to have necessary cervical screen themselves; people who take care of them, either families or healthcare professional in facilities need to help them to have cervical screen timely and appropriately.

Next we extract important non-code features

	index	feature	weight	screen_rate	frequency	description
5151	56116597	14015592	0.122118	0.937310	0.001984	primary_practitioner_id
159	12255777	29416393	-0.102089	0.244323	0.001902	primary_practitioner_id
483	265094781	16444459	0.261919	0.977121	0.001812	primary_practitioner_id
2375	262447155	14007032	0.139398	0.645636	0.001643	primary_practitioner_id
4688	241647461	12945837	0.123303	0.631748	0.001616	primary_practitioner_id

The only features that the algorithm found useful in diagnosis_head.csv that are not diagnosis_code are primary_practitioner_ids.



The histogram plot shows that more than 2000 practitioner ids correspond to a screener rate greater than 0.9. However, most of them are very infrequent. The scatter plot on the right shows that there are three practitioner IDs that correspond to screener rate greater than 0.9 and frequency greater than 0.1%, which are "16444459", "12709441", and "14015592". So who are these practitioners?

Insight 6: Some practitioners' job directly relates to screening or cervical cancer.

	physician_id	practitioner_id	state	specialty_code	specialty_description	CBSA
109453	12709441	12709441	CA	DR	DIAGNOSTIC RADIOLOGY	41860
430336	14015592	14015592	CA	DR	DIAGNOSTIC RADIOLOGY	41860
677594	13116555	16444459	CA	OBG	OBSTETRICS AND GYNECOLOGY	41860

Apparently their jobs are strongly related to screening. Interesting thing is they are all from the same location of California.

Insight 7: Nurse's experience matters

	index	feature	weight	screen_rate	frequency	description
8	240540920	GYN	0.050469	0.714609	0.116578	specialty_code
6	202466831	DMP	0.052519	0.649283	0.039235	specialty_code
7	28074032	NMW	0.076270	0.812271	0.024508	specialty_code
0	90934848	SP	0.055063	0.664782	0.018182	specialty_code
3	16614087	FOP	0.054828	0.660666	0.001583	specialty_code
5	103656617	PLM	-0.062786	0.476569	0.000977	specialty_code
1	118656641	PCH	0.073176	0.635379	0.000239	specialty_code
4	124656680	PYA	-0.075508	0.619048	0.000054	specialty_code
2	170165281	SNP	-0.059112	0.250000	0.000014	specialty_code

The only features that the algorithm found useful in physicians.csv is specialty code. The table above shows that the specialty code NMW, NURSE MIDWIFE, correspond to 81% screener rate and 2.4% frequency which is the most useful indicator for screeners. The specialty code SNP, SCHOOL NURSE PRACTITIONER, strongly indicate non-screeners but it is not frequent. A simple googling shows that NMW is certified practitioners while SNP is obviously less experienced, which makes intuitive sense.

Analyze the extracted features' predictive power

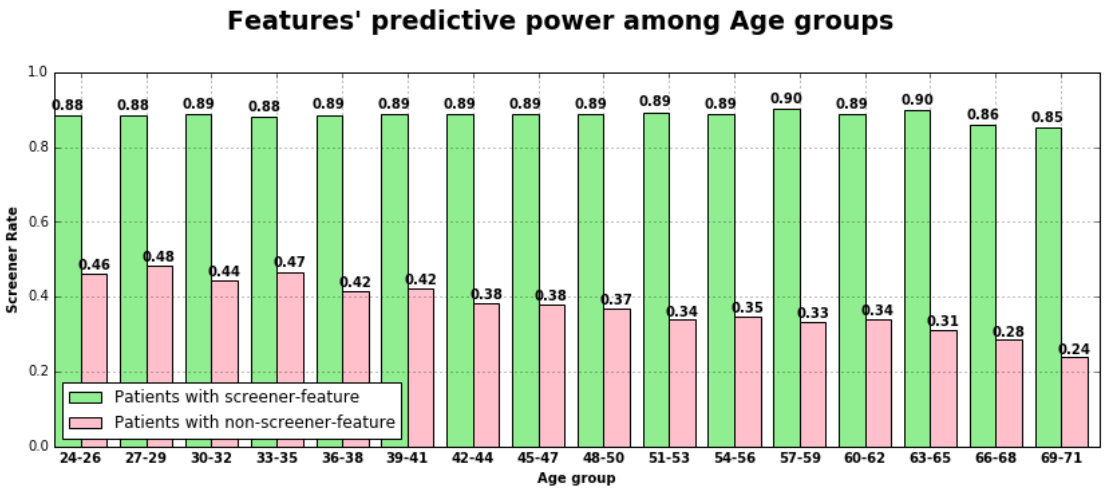
```
(95, 10)
```

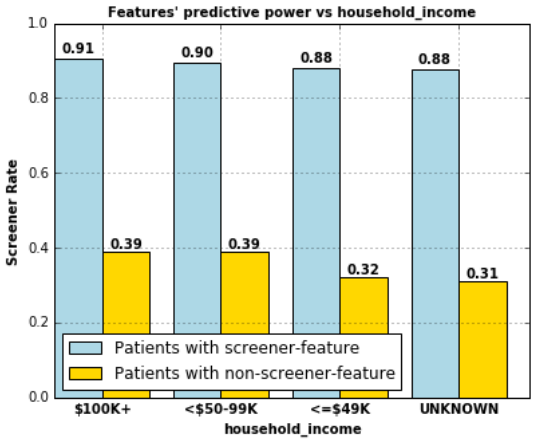
We further extract 95 most important features from our model and study their predictive power with different factors like age, ethinicity, education and income level.

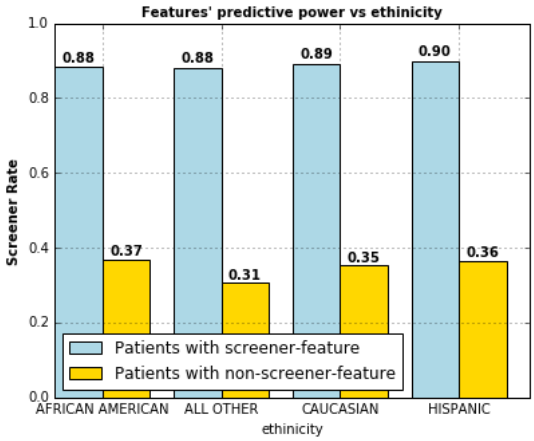
```
(229782, 103) (20187, 103) (1157817, 103) (0, 103)
0.885674247765 0.342745331154 0.557785038568
```

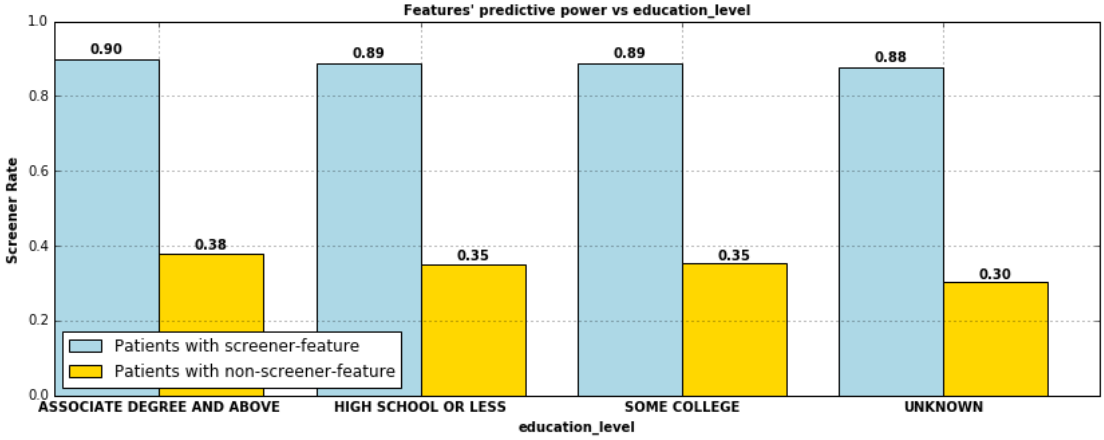
To make it simple, we define patients with screener-features as the ones who has at leat one non-zero screener feature and no non-screener feature.

```
24084
```




Features' predictive power vs household_income
Screener Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.91
0.39
0.90
0.39
0.88
0.32
0.88
0.31
Patients with screener-feature
Patients with non-screener-feature
$100K+
<$50-99K
<=$49K
UNKNOWN
household_income


Features' predictive power vs ethinicity
Screener Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.88
0.37
0.88
0.31
0.89
0.35
0.90
0.36
Patients with screener-feature
Patients with non-screener-feature
AFRICAN AMERICAN
ALL OTHER
CAUCASIAN
HISPANIC
ethinicity


Features' predictive power vs education_level
Screener Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.90
0.38
0.89
0.35
0.89
0.35
0.88
0.30
Patients with screener-feature
Patients with non-screener-feature
ASSOCIATE DEGREE AND ABOVE
HIGH SCHOOL OR LESS
SOME COLLEGE
UNKNOWN
education_level