

مدل پیش بینی پذیرش دانشجویان در مقاطع تحصیلات تکمیلی دانشگاه های خارج از کشور

> دانشجو: شهریار یزد*ي* پور

استاد راهنما: دکتر ناهید طاهریان



- مقدمه و هدف پروژه
- بررسی و مطالعه تحقیقات پیشین
- جمع آوری داده و چالش های آن
  - مراحل تحلیل داده
  - بررسی و پیش پردازش داده ها
    - داده کاوی روی داده ها
      - نتایج

فهرست

# انگیزه

- بسیاري از دانشجویان مشتاق به ادامه ي تحصیلات آکادمیک خود در کشور هاي پیشرفته ترند
  - هزینه ی بالایِ مالی و زمانی اقدام برای گرفتن پذیرش
- مشكل: ذهن انسان توانايى نگهداري و تحليل همزمان حجم زيادي از نمونه ها با خصيصه هاي متفاوت را ندارد.

# طراحی مدلی برای پیش بینی:

- علم داده کاوی
- داده های پیشین

**داده كاوي** شامل استفاده از ابزار هاي پيشرفته تحليل داده به منظور كشف الگوهاي معتبر و روابط در مجموعه داده هاي بزرگ است. اين ابزار ها، مدل هاي آماري، الگوريتم هاي رياضي و متد هاي يادگيري ماشين مي باشند.



## بررسى و مطالعه تحقیقات پیشین

 A Review on Recommender Systems for University Admissions (2015)

Aditi Bhide, Karan Karnik, Pradnya Badge, Varun Joshi, Vina M. Lomte

Paper ID: ART20163309

Data From schools portal

C4.5 Classifier + Feedback

 An Automated University Admission Recommender System for Secondary School Students (2009)

Simon Fong and Robert P. Biuk-Aghai

Paper Code: 26960-IJAES

Input: CGA, Age, Birth Place, GRA\_TA

Recommended universities: University of Macau , Taiwan , China , In Asia

Main work: Customized C4.5 Classifier

# A Decision Tree Classification Model for University Admission System

(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 10, 2012

Attribute	Possible values					
Gender	Student's gender					
	Male					
	Female					
HS_Type	Type of high school study					
	TS = Scientific Study					
	TL = Literature Study					
	<ul> <li>TU = Unknown/Missing</li> </ul>					
HS_Grade	High school grade					
	<ul> <li>A = mark ≥ 85</li> </ul>					
	<ul> <li>B = 75 ≥ mark &gt; 85</li> </ul>					
	<ul> <li>C = 65 ≥ mark &gt; 75</li> </ul>					
	<ul> <li>D = 50 ≥ mark &gt; 65</li> </ul>					
Area	Code for student's region city (116					
	distinct value)					

Accuracy: 65%

**20K Records** 

- Discovering Knowledge regarding Academic Profile of students pursing Graduate studies in World's Top Universities
- Graduate School Recommender System: Assisting Admission Seekers to Apply for Graduate Studies in Appropriate Graduate Schools

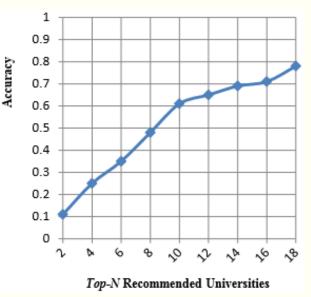
December 2016, Dhaka, Bangladesh

By Mahamudul Hasan, Shibbir Ahmed, Deen Md. Abdullah, and Md. Shamimur Rahman

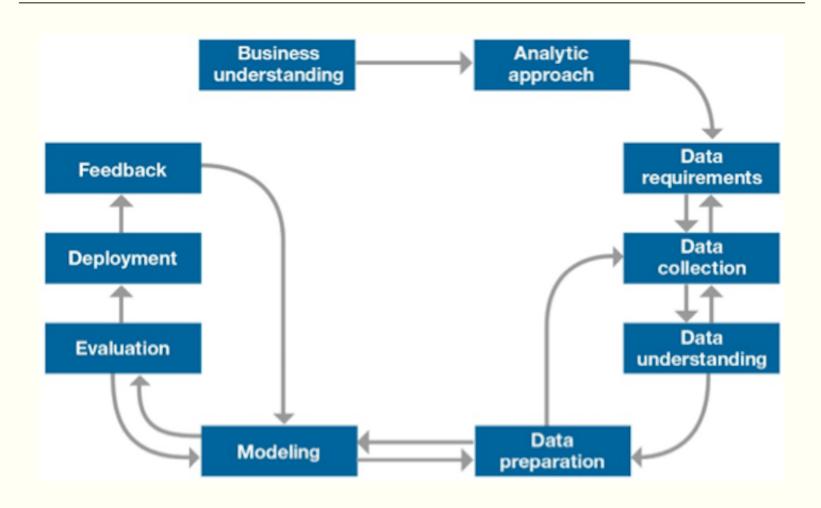
## بررسی و مطالعه تحقیقات پیشین

- Focus on PhD
- Focus on applying for USA universities
- Accuracy:
  - 90.342% (20 recommendation)
  - 50% (8 recommendation)

	Undergrad Department		
	Undergrad University		
	Undergraduate CGPA		
Academic Data	Research Area		
	Research Experience		
	Job Experience		
	Publication Records		
	GRE (Verbal, Quantitative, Analytical		
	Writing, Total Score)		
Standardized Test Score Data	TOEFL/ IELTS (Reading, Listening, Speaking, Writing, Total Score)		
Personal Data	Gender		



#### Methodology



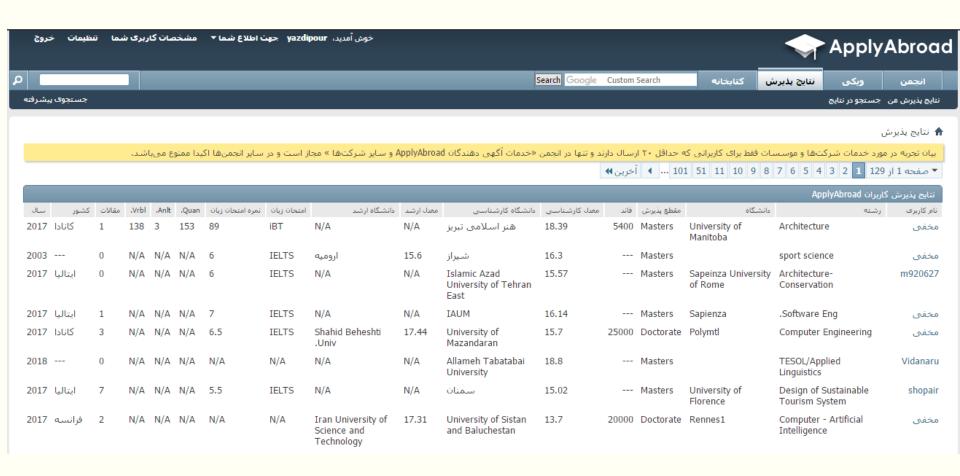
# DATA REQUIREMENT



# اولین انجمن تحصیلات تکمیلی در خارج از کشور

اپلای ابرود ApplyAbroad.org مجموعهای شامل اطلاعات متنوع در زمینه گرفتن پذیرش از دانشگاههای خارج از ایران، مراحل و اطلاعات لازم جهت دریافت ویزای تحصیلی و گرفتن مهاجرت از کشورهای مختلف، از جمله کانادا، استرالیا و نیوزیلند است و با هدف کمک به دانشجویان ایرانی برای ادامه تحصیل در خارج از کشور ایجاد شده است. در حال حاضر فروم و ویکی بزرگترین مجموعه اطلاعات اپلای ابرود یه شمار می آیند که شامل اطلاعات زیادی میباشند.

#### www.applyabroad.org/forum/admissionreport.php



http://www.applyabroad.org/forum/misc.php?do=showrules

قانون کی) در صورتی که میخواهید از مباحث و آموزشهایی که در این انجمنها ذکر شده است در جایی دیگر و یا در سایت خود استفاده کنید ذکر نام سایت و نام نویسنده مطلب و لینک مطلب الزامیست.

# http://www.applyabroad.org/forum/admissionreport.php More than 4000 records in Table

															کاربران ApplyAbroad
سال	كشور	مقالات	.Vrbl	.Anlt	.Quan	نمره امتحان زبان	امتحان زبان	دانشگاه ارشد	معدل ارشاد	دانشگاه کارشناسی	معدل كارشناسي	فاند	مقطع پذیرش	دانشگاه	رشته
2017	ایتالیا	0	N/A	N/A	N/A	6	IELTS	N/A	N/A	Islamic Azad University of Tehran East	15.57		Masters	Sapeinza University of Rome	Architecture- Conservation
2017	ايتاليا	1	N/A	N/A	N/A	7	IELTS	N/A	N/A	IAUM	16.14		Masters	Sapienza	.Software Eng
2017	كانادا	3	N/A	N/A	N/A	6.5	IELTS	Shahid Beheshti .Univ	17.44	University of Mazandaran	15.7	25000	Doctorate	Polymtl	Computer Engineering
2018		0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Allameh Tabatabai University	18.8		Masters		TESOL/Applied Linguistics
2017	ايتاليا	7	N/A	N/A	N/A	5.5	IELTS	N/A	N/A	سمنان	15.02		Masters	University of Florence	Design of Sustainable Tourism System
2017	فرانسه	2	N/A	N/A	N/A	N/A	N/A	Iran University of Science and Technology	17.31	University of Sistan and Baluchestan	13.7	20000	Doctorate	Rennes1	Computer - Artificial Intelligence
2017	سوئد	9	152	5	159	100	iBT	Azad, Science Research	17.43	Azad, Science Research	14.7		Doctorate	KTH	Biomedical Engineering
2017	آمریکا 110	0	150	4	160	112	iBT	N/A	N/A	azad	16.34	14000	Masters	ODU	mechanical engineering

# http://www.applyabroad.org/forum/admissionreport.php More than 4200 records in Forms

	r0a0m0k نتایج پذیرش
Poultry Nutrition مقطع پذیرش: Poctorate	رشته:
معدل: 15.48 دانشگاه: Ferdowsi University of Mashhad رشته: animal science	کارشناسی:
معدل: 18.08 دانشگاه: Ferdowsi University of Mashhad رشته: Poultry Nutrition	کارشناسی ارشد:
90 نوع امتحان زبان: iBT	نمره امتحان زبان:
158 : Qunt. 3 : Anlt. 148 : Vrbl.	نمره امتحان GRE:
خارجى: 4 داخلى: 2	تعداد مقالات:
2017	ساك شروع تحصيل:
University of Guelph	دانشگاههای پذیرفته شده:
University of Guelph کشـور: کانادا	دانشگاه انتخاب شده:
40000   نوع فاند: Award	ساپورت مالی در ساك:
استانبول	محل دریافت روادید:



# تخمیمی از حجم داده ها در روشهای مختلف گرفتن داده ها

Type of Extracting	Avg Size Per Page (4200pages)	Avg Size of whole pages		
Browser	700Kbytes	700*4200/1024=2870Mbytes		
Crawler (no reference)	30Kbytes	30*4200/1024=123Mbytes		
Crawler + Filter (array of strings)	0.7KBytes	0.7*4200/1024=2.8MBytes		
Crawler + Filter (array of objects)	0.3Kbytes	0.3*4200/1024=1.2MBytes		

حجم بعد از پروسه ی گرفتن اطلاعات: 1.5MB

# Library



https://curl.haxx.se/



http://simplehtmldom.sourceforge.net/

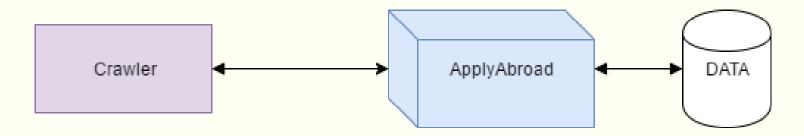
# بدست آوردن داده ها و چالش های آن

#### ■ دانلود صفحات

http://www.applyabroad.org/forum/admissionreport.php?do=viewreport&ar=<X>

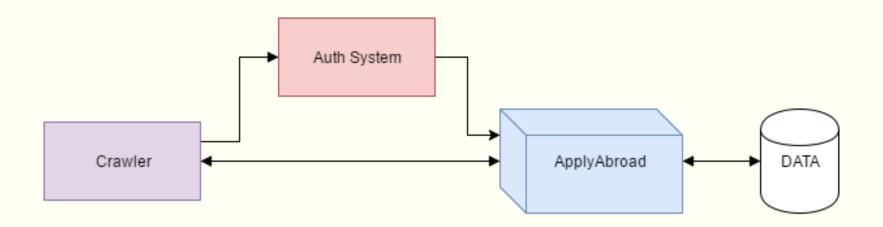
X: From 1 to 4275

- پارس کردن صفحه وب ( با کمک کتابخانه SimpleHtmlDOM)
  - استفاده از Regex برای طبقه بندی داده ها
    - تبدیل به فرمت JSON
  - چالش : عبور از اعتبارسنج + به دام افتادن crawler



# بدست آوردن داده ها و چالش های آن

- عبور از اعتبار سنجی سایت با کمک کتابخانه ی
  - سایت از VBulletin Forum CMS/استفاده میکند

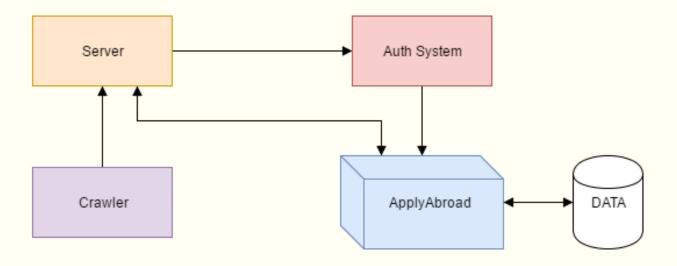


#### Sources (Auth Bypass)

```
19 $ch = curl init();
20 curl setopt($ch, CURLOPT HEADER, false);
21 curl setopt($ch, CURLOPT COOKIEFILE, 'veri.js');
22 curl setopt($ch, CURLOPT COOKIEJAR, 'veri.js');
    curl setopt($ch, CURLOPT RETURNTRANSFER,true);
    curl setopt($ch, CURLOPT COOKIESESSION,true);
25 curl setopt($ch, CURLOPT SSL VERIFYPEER,false);
26 curl_setopt($ch, CURLOPT_SSL_VERIFYHOST, false);
27 curl setopt($ch, CURLOPT FOLLOWLOCATION,true);
28 curl setopt($ch, CURLOPT USERAGENT, $ SERVER['HTTP USER AGENT']);
    curl setopt($ch, CURLOPT REFERER, $url.'index.php');
30 curl setopt($ch, CURLOPT URL, $url.'login.php?do=login');
31 curl setopt($ch, CURLOPT POST, 1);
32 curl_setopt($ch, CURLOPT_POSTFIELDS, "vb_login_username=$username
        &vb login password&s=&securitytoken=guest&do=login&vb login md5password=".md5($password).
        &vb login md5password_utf=".md5($password));
   $exec = curl exec($ch);
34 curl_setopt($ch, CURLOPT_POST, 0);
35 curl setopt($ch, CURLOPT REFERER, $url.'login.php?do=login');
   curl_setopt($ch, CURLOPT_URL, $url.'clientscript/vbulletin global.js?v=373');
   $exec = curl exec($ch);
37
38 .
39
40
```

# بدست آوردن داده ها و چالش های آن

- استفاده از وب سرویس رایگان (000webhost.com)
  - تاثیر مستقیم روی سرعت بخشیدن
    - درگیر نشدن سیستم های شخصی
  - دانلود 1.5 مگابایت اطلاعات بجای 120 مگابایت



# DATA UNDERSTANDING DATA PREPROCESSING

#### 3736 RECORDS (FEB 2017)

```
"3": {
                                           "papersGLOB": "0",
    "field": "Civil Engineering /
Transportation",
                                            "papersIRAN": "8",
    "apDegree": "Doctorate",
                                            "year": "2010",
    "bachelorsAvg": "15.7",
                                            "accUni": "EPFL Switzerland,
    "bachelorsUni": "فردوسي مشهد",
                                       Northwestern",
    "masterAvg": "19",
                                            "rejUni": "UC Berkeley, Georgia
                                       Tech",
    "masterUni": "Portland State
                                            "apUni": "Northwestern",
University",
    "engExamResult": "100",
                                            "country": "آمریکا",
    "engExamType": "iBT",
                                           "fund": "63000",
    "engExamQuan": "760",
                                            "fundType": "RA"
    "engExamAnlt": "3",
                                         }
    "engExamVrbl": "330",
```

# بررسی و پیش پردازش داده ها

	accUni	apDegree	apUni	bachelorsAvg	bachelorsUni	country	engExamAnlt	engExamQuan	engExamResu
2	n/a	Doctorate	Heriot Watt University, Edinburgh	17.74	أزاد اسائمی امیدیه	ادگلستان	n/a	n/a	n/a
3	EPFL Switzerland, Northwestern	l	Northwestern	15.70	فر دوسبی مشهد	أمريكا	3	760	100
5	n/a	Doctorate	Cornell	17.68	امیر کبیر	أمريكا	3.5	800	100
6	Dalhousie, Clemson, TUDelft, NTU-NUS (SINGA sc	Masters	Chalmers	16.00	صنعتى اميركبير	سوكد	3	800	92
7	Carlos III Madrid, FHTW Berlin, Chalmers, Poli	Masters	Carlos III Madrid, Spain	7.67	پلی تکنیک بھارست، رومانی معدل 7.67 از 10	اسپانیا	n/a	n/a	103

# بررسی و پیش پردازش داده ها

Degree

Year

**Country** 

**TOFEL / IELTS** 

**GRE** 

**Fund** 

**GPA** 

**Papers** 

**Universities** 

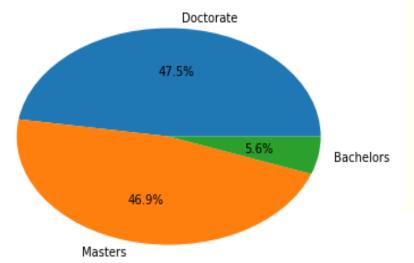
# Library



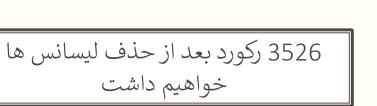


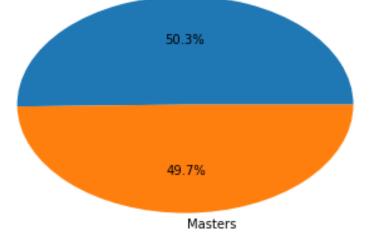


# Degree



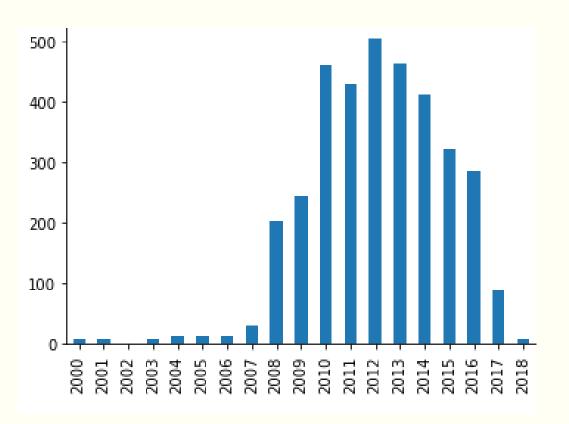
Degree	Count
Doctorate	1773
Masters	1753
Bachelors	210





Doctorate

#### Year



- چالشها:
- چندین داده به صورت سال شمسی + تبدیل انها به میلادی
  - 13 داده گم شده + جایگزینی انها با داده بعدی رکود

### **Country**

US 1137	Belgium 35
Canada 593	France 33
Germany 234	New Zealand 29
Australia 185	Swaziland 28
Sweden 161	Spain 24
Italy 146	Singapore 17
England 68	Austria 17
Finland 68	China 8
Malaysia 65	India 7
Turkey 44	Russia 3
Netherlands 43	Czech Republic 2
Norway 41	Japan 2

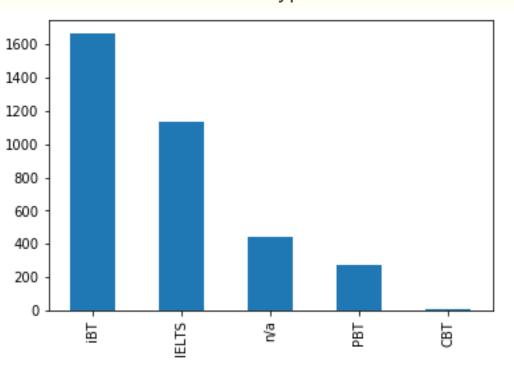
- تبدیل اسامی فارسی کشور ها به انگلیسی
- 536 داده ی گم شده در مرحله ی اول
- بدست آوردن کشور از نام دانشگاه پذیرفته شده بعد از کامل کردن لیست دانشگاه ها
  - کاهش 536 داده ی گم شده به 276 داده در مرحله ی دوم

# Encoding

Classified	Countries
0	['united states']
1	['canada','australia','switzerland']
2	['sweden','finland','norway','iceland','denmark']
3	<pre>['germany','france','austria','poland','spain','italy']</pre>
4	<pre>['united kingdom','netherlands','belgium','new zealand']</pre>
5	['singapore','korea, republic of','japan']
6	<pre>['malaysia','brazil','iran, islamic republic of','south africa','czech republic','turkey','russian federation']</pre>
7	<pre>['china','india','pakistan','albania','ireland','mexico','phi lippines','cyprus']</pre>
8	Others

# TOFEL / IELTS





442 داده گم شده + جایگذاری با میانگین

# **TOFEL / IELTS**

طبق مطالب بیان شده در سایت زیر همه ی نتایج زبان را به IELTS تبدیل میکنیم http://www.eurogates.nl/en-TOEFL-IELTS-score-conversion/

OEIC	TOEFL Paper	TOEFL CBT	TOEFL IBT	IELTS	VEC Online Score	Approximate VEC Level
0 - 250	0 - 310	0 - 30	0 - 8	0 - 1	0 - 34	Beginner
0 - 230	310 - 343	33 - 60	9 - 18	1 - 1.5	35 - 38	Middle beginner
255 - 400	347 - 393	63 - 90	19 - 29	2 - 2.5	39 - 45	Upper Beginner
255 - 400	397 - 433	93 - 120	30 - 40	3 - 3.5	46 - 53	Low Intermediate
405 - 600	437 - 473	123 - 150	41 - 52	4	54 - 57	Middle Intermediate
403 - 600	477 - 510	153 - 180	53 - 64	4.5 - 5	58 - 65	Intermediate
605 - 780	513 - 547	183 - 210	65 - 78	5.5 - 6	66 - 73	Low advanced
605 - 760	550 - 587	213 - 240	79 - 95	6.5 - 7	74 - 81	Middle advanced
785 - 900	590 - 637	243 - 270	96 - 110	7.5 - 8	82 - 90	Advanced
905 - 990	640 - 677	273 - 300	111 - 120	8.5 - 9	91 - 100	Upper Advanced

# **TOFEL / IELTS**

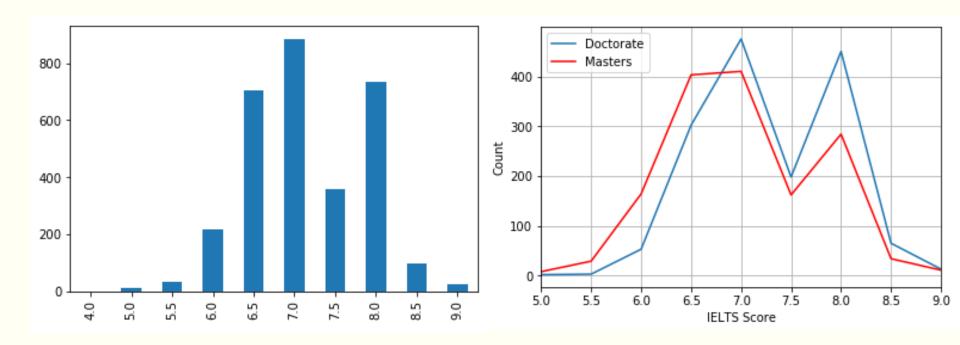


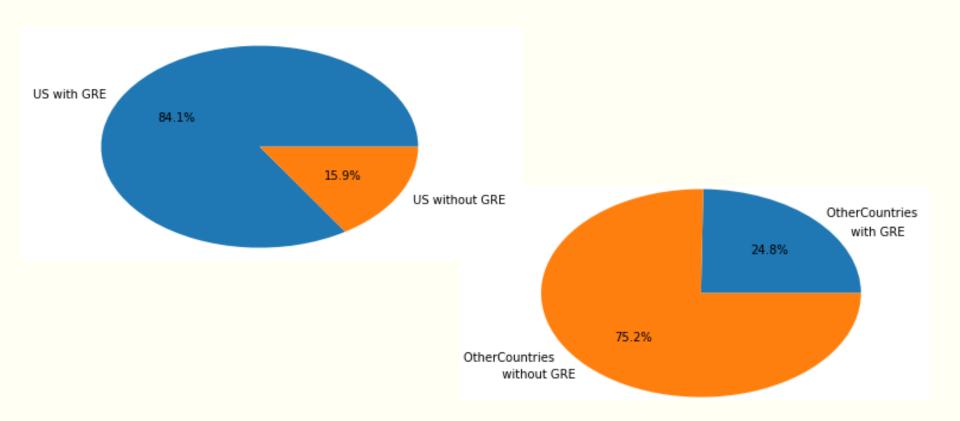
TABLE VI. DISCRETIZATION FORMAT TABLE FOR IELTS SCORES

	Range of Scores (S)						
Classified Name	Reading/Listening/Speaking/Writing	Total					
VeryHigh	S > 7.5	S > 7.5					
High	7.0 ≤ S ≤ 7.5	$7.0 \le S \le 7.5$					
Medium	$6.0 \le S \le 6.5$	$6.0 \le S \le 6.5$					
VeryLow	5 ≤ S ≤ 5.5	5 ≤ S ≤ 5.5					
Low	S <5	S <5					

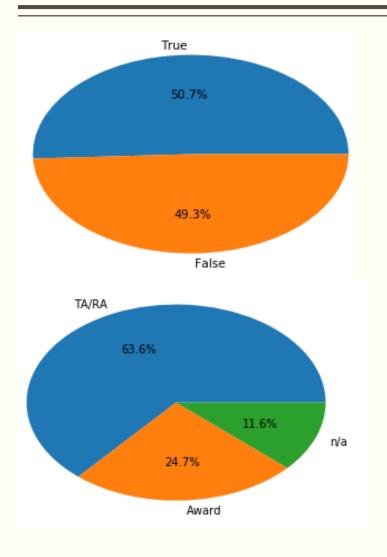
<sup>\*</sup> Based on : Discovering Knowledge regarding Academic Profile of students pursing Graduate studies in World's Top Universities

### **GRE**

1560 داده ی از امتحان GREداریم. این نوع امتحان پیش نیاز برای پذیرش در دانشگاه های امریکایی است



### **Fund**



50.7% از دانشجویان موفق به دریافت کمک هزینه شده اند

Туре	Count
TA/RA	1138
Award	442
n/a	208

### **GPA** (grade point average)

حرف یا عددی است که در آموزشگاهها یا دانشگاهها پس از پرسیدن درس یا انجام امتحانی با آن کمیت، دانشِ دانشجویان را میسنجند.

Ref: Wikipedia.com

### **Grading Scale:**

Grade	Scale
Α+	18 - 20
А	16 - 17.99
В	14 - 15.99
С	12 - 13.99
D	10 - 11.99
F	0 - 9.99

### Base on:

http://www.foreigncredits.com/resources/gpa-calculator/iran

### **GPA**

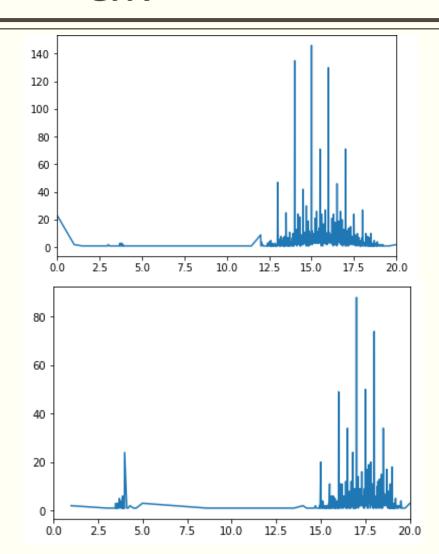
(10, 20] 3454

```
(0, 4]
               36
(-14, 0]
               23
(20, 18050]
            10
(4, 10]
Name: bachelorsAvg,
   dtype: int64
(10, 20] 1688
(0, 4]
            110
(20, 1820]
          24
(4, 10]
 Name: masterAvg,
  dtype: int64
```

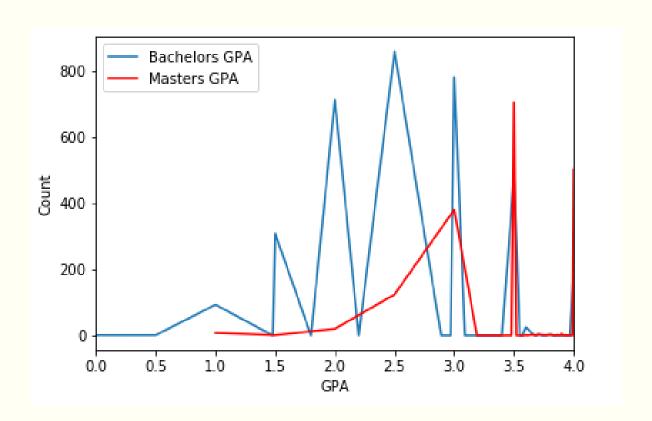
### **GPA**

Bachelors GPA

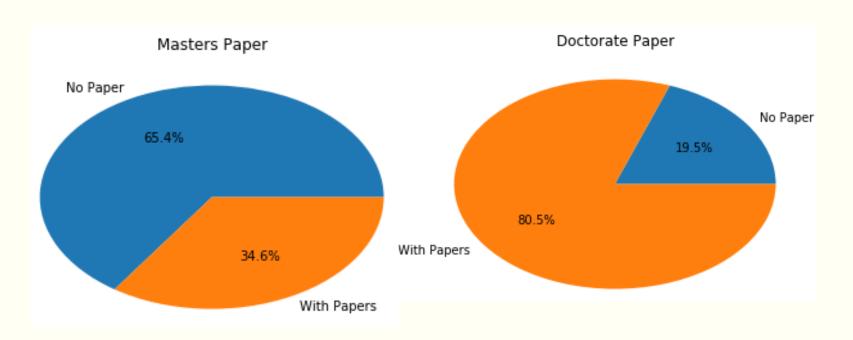
Masters GPA



### **GPA**



### **Papers**



mean 2.760068

std 4.594755

min 0.000000

25% 0.000000

50% 1.000000

75% 4.000000

max 56.000000

### **Fields**

- 1519 مدل رشته به صورت خام
- 1307 مدل داده بعد از حرف کوچک کردن داده ها
- 655 مدل داده بعد از فیلتر کردن به کمک کلید واژه ها
- 58 مدل داده بعد استفاده از کتابخانه ی difflib روی داده های استفاده شده
  - کاهش به 20 گروه

### Library



https://docs.python.org/2/library/difflib.html

- از الگوریتم "gestalt pattern matching" استفاده میکند
- ایده: بزرگترین برداری که طبق الگوریتم رگرسیون Ratcliff and ایده: بزرگترین برداری که طبق الگوریتم رگرسیون Obershelp
  - $n^4$  پیچیدگی زمانی در بدترین حالت •

### **Universities**

دانشگاه پذیرفته شده دانشگاه لیسانس دانشگاه فوق لیسانس



- 10K list of world universities
  - https://github.com/endSly/world-universities-csv
- List of all countries with their 2 digit codes:
  - http://data.okfn.org/data/core/country-list
- 4K universities name CWUR Ranking:
  - <a href="https://github.com/shubhanerur/RankingUniversities">https://github.com/shubhanerur/RankingUniversities</a>
- Iranian Universities List:
  - https://fa.wikipedia.org/wiki/رتبهبندی\_دانشگاههای\_ایران

### Universities

112	تهران	elmo sanat 1
84	شریف	از اد رامسر
azad	82	private university 1
university of tehran	80	دانشگاه تفرش (سر اسري)
76	علم و صنعت	امير کيبر
71	امیرکبیر	غیر انتفاعی بهمنیار کر مان
61	صنعتی اصفهان	azad of takestan 1
amirkabir	46	islamic azad university of khorasgan(isfahan) 1
45	تبريز	shiraz u 1
iust	45	از اد خرم اباد
sharif	44	damgan azad university 1
43	شیراز	art university of isfahan 1
41	فر دو سی مشهد	موسسه آموزش عالى سجاد
39	دانشگاه تهران	امام حسین
sharif university of ted	chnology 38	sadjad institute of higher ed 1
36	خواجه نصير	دانشگاه صنعتی امیرکبیر
shiraz university	32	جهرم
n/a	32	مرکز تحصیلات تکمیلی در علوم پایه زنجان
32	صنعتى شريف	دانشگاه آز اد اسلامی تهران شمال
32	صنعتى امير كبير	دانشگاه آز اد و احد تبریز
31	شهید بهشتی	روزبه زنجان
university of tabriz	29	علامه
iau	29	منهدسی نرم افز ار
28	اصفهان	azad - tehran north 1
azad university	28	زاهدان
tehran	27	azad uni - tehran medical branch 1
islamic azad universit	y 26	علم و صنعت ايران ( واحد اراک)
ferdowsi university of	mashhad 22	kharazmi university 1
shahid beheshti	20	راغب اصفهاني
17	گیلان	kar university 1

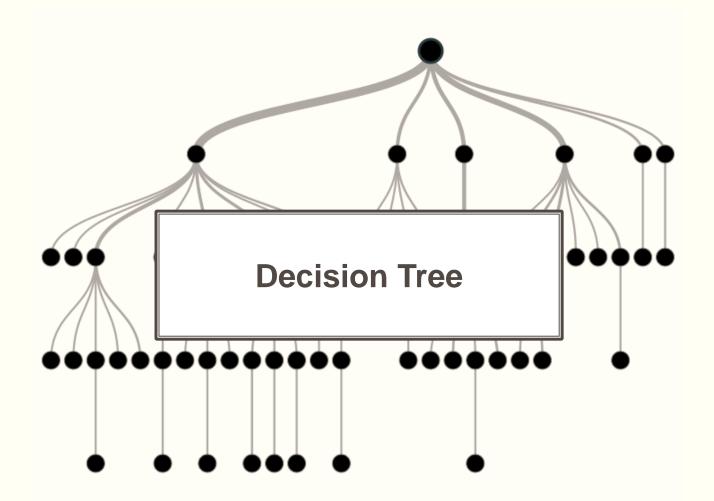
### **Universities**

n/a	426	university of saskatchewan	
politecnico di milano	46	15	
kth	46	university of manitoba	15
university of alberta	32	university of houston	13
concordia	32	university of victoria	13
utm	27	calgary	13
ubc	26	university of waterloo	13
ntnu	24	upm	13
unsw	23	arizona state university	12
concordia university	22	ncsu	12
politecnico di torino	22	•••	
university of calgary	21	university of leoben	1
rwth aachen	20	imperial college london	1
mcgill	19	university of texas	1
university of toronto	19	karlsruher institut fur technologie	1
chalmers	19	dalhousoie university	1
alberta	19	universite de montreal	1
sfu	18	mapua	1
waterloo	18		
mcmaster	16		
virginia tech	16		
3	. •		

# MODELING **EVALUATION**

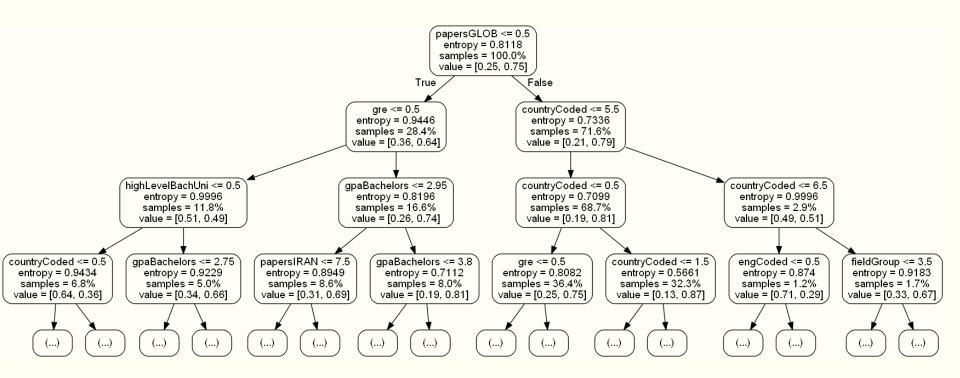
### Library





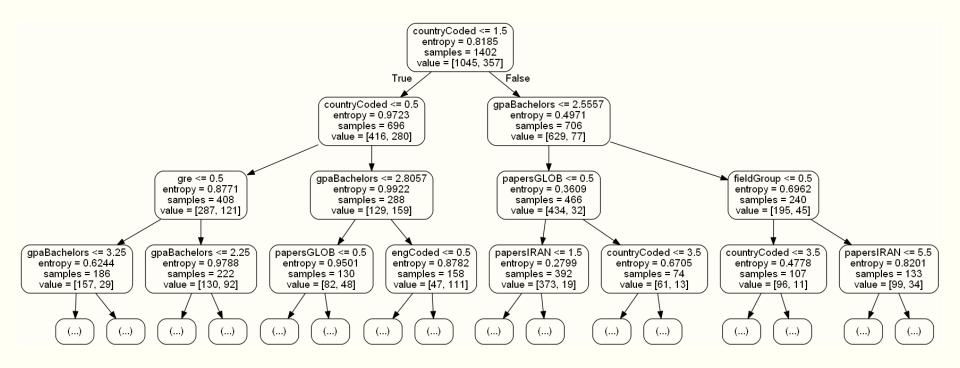
### **Decision Tree**

Getting scholarships for PhD Best Accuracy: **80.0%** in **5**<sup>th</sup> **Level** 



### **Decision Tree**

Getting scholarships for MS Best Accuracy : **80.6%** in **5**<sup>th</sup> **Level** 



### **Decision Tree Sample**

### Input

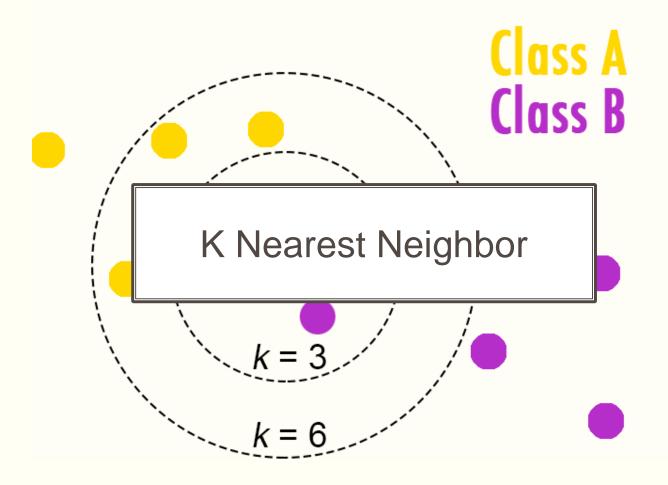
Country Coded	Eng. Coded	Field Group	GPA Bachelors	-   ( <del>-</del>		Papers GLOB	Papers IRAN
1	2	0	2.5	False	False	1	0

### Output

```
clf.predict([me])
array([ True], dtype=bool)

clf.classes_
array([False, True], dtype=bool)

clf.predict_proba([me])
array([[ 0.36111111,  0.63888889]])
```



### **Supervised Classification**

### **K-Nearest**

Finding best Country For 48 distance algorithm Best Accuracy: about 31%

### **K-Nearest**

Getting Fund for MS
For 48 distance algorithm
Best Accuracy: 76.89% With City Block Distance

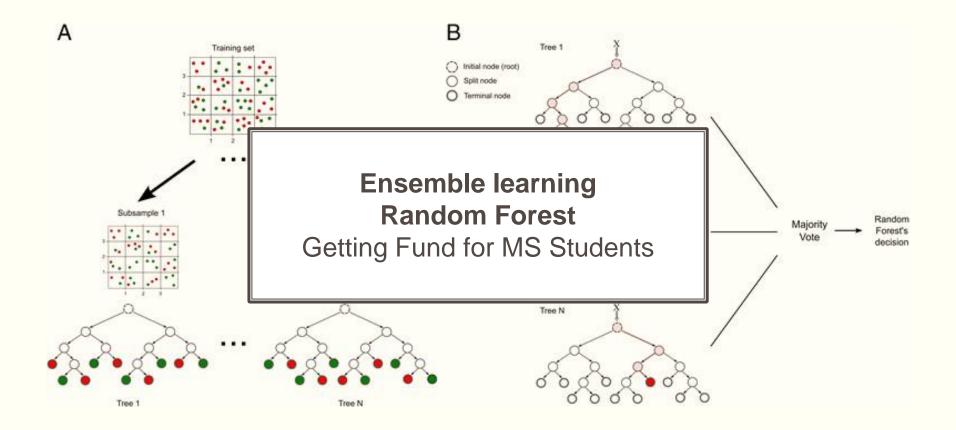
### K Nearest getting fund sample

### Input

Country Coded	Eng. Coded	Field Group	I (Free I		High Level bach Uni	Papers GLOB	Papers IRAN
1	2	0	2.5	False	False	1	0

### Output

Country Coded	Eng. Coded	Field Group	GPA Bachelors	Gre	High Level bach Uni	Papers GLOB	Papers IRAN
1	1	2	2	False	True	4	2
5	1	0	3	False	True	1	3
2	1	0	2	False	False	1	0



### Random Forest Test

```
In [92]: me=[1, 2, 0, 2.5, False,False,0,0]
         clf.predict([me])
Out[92]: array([False], dtype=bool)
In [93]: clf.predict_proba([me])
Out[93]: array([[ 0.8, 0.2]])
In [94]: clf.classes
Out[94]: array([False, True], dtype=bool)
In [95]: me=[1, 2, 0, 2.5, False,False,1,0]
         clf.predict([me])
Out[95]: array([ True], dtype=bool)
         clf.predict_proba([me])
Out[96]: array([[ 0.4, 0.6]])
```

## **K-Means Clustering Unsupervised learning** Recommending Universities

### KNN Test – 10 Clusters

```
# 'countryCoded', 'engCoded', 'fieldGroup', 'gpaBachelors', 'gre',
# 'highLevelBachUni', 'papersGLOB', 'papersIRAN', 'fund'
me=[1,2, 0,2.5,False,False,1,0,True]
pred=model.predict([me])
olddf.ix[cp[cp.label==pred[0]].index]['targetUni'].value counts()
concordia
                    26
simon fraser
                    18
calgary
                    18
alberta
                    17
manitoba
                    14
melbourne
                    14
british columbia
                    13
western ontario
                    10
toronto
waterloo
                     9
```

### KNN Test – 2-20 Clusters

### Input

Country Coded	Eng. Coded	Field Group	GPA Bachelor s	Gre	High Level bach Uni	Papers GLOB	Papers IRAN	Fund
1	2	0	2.5	False	False	1	0	True

349
223
207
205
180
177
134
128
119
118
101
89
86
85
83

### برنامه زمانی پروژه

قدم چهارم قدم سوم قدم دوم قدم اول پروپوزال اوپن سورس بدست آوردن داده ها بررسی تکنیکهای داده کاوی هدف گذاری بهینه سازی ساختار دادن به داده مشخص کردن پور تال منابع Recommender System مشخص کردن پیش پردازش ابزارها



### THIS PROJECT IS GOING TO BE OPEN SOURCE.

### Tools

















با تشکر

خرداد – نود و شش شمسی