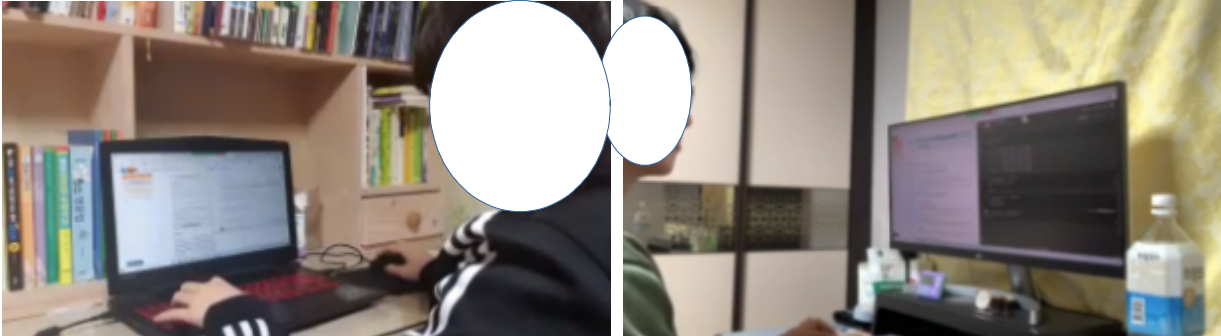


## 2021.기계학습.퀴즈2.시험지 - 2021년 6월 9일 (수)

### [ 온라인 시험 바른 자세 예시 ]

아래와 같이 바른 자세 예시를 따르지 않는 학생은 퇴실 조치 합니다.

자세 검사를 마친 학생에 한하여 시험지의 비밀번호가 제공됩니다.



### ## 온라인 시험 주의 사항

- 시험 중 화장실에 다녀올 수 없음
- 화면 녹화 본 제출 필수 (대용량 이메일), 화면 녹화 중단시 녹화된 곳 까지만 점수 인정
- 화면 녹화는 반드시 7시부터 시작, 녹화 시작시 컴퓨터 하단에 그 어떠한 창도 실행되어 있으면 안됨
- 코드 제출 필수(구글드라이브), 반드시 제출 코드로 리더보드 재현이 가능해야 인정
- OJ를 통한 IP 필터링을 통해 모여서 시험을 보는 행위 방지함, 이에 테더링 사용 불가
- 고사 중 부적절한 자세로 인한 경고 3회 이상 누적시 시험 0점 처리
- 듀얼모니터 사용 불가, 시험 시작 후 듀얼 모니터 사용 중이 아님을 입증하는 화면 캡처 제출 필수
- 시험은 오픈 메뉴얼 북 (메뉴얼 안에서 검색 가능)만 허용, 강의안과 책 불가, 인터넷 검색 불가
- 고시 시작 두 시간 이후 퇴실 가능하나 (퇴실 시 반드시 퇴실 하겠다고 알릴 것) 재 입실 불가
- 시험 중 캐글 리더보드 여전히 20회 제출 제한 있음
- 실시간 중계되는 감독 화면은 녹화되어 보관됨
- 고사 중 공지 및 질문을 위해서는 슬랙을 사용할 예정, 반드시 본인 PC에 슬랙 설치 필수
- 고사 중 카카오톡과 같은 메신저는 반드시 종료할 것. 적발시 F 처리
- 기존 코드를 참고하는 행위 절대 불가, 적발시 F
- 대리 시험 절대 불가, 적발시 학사경고

### ## 제출

- 화면 녹화 파일 제출: [admin@rcv.sejong.ac.kr](mailto:admin@rcv.sejong.ac.kr)
  - : 화면 녹화 본 중 듀얼모니터 부분 화면 캡처 제출: 동영상과 함께 이메일로 제출
- 코드 및 기타 파일
  - : 공유된 구글 드라이브

퀴즈2는 총 5문제로 총 3점이 할당되어 있으며, 문제당 0.6점이 할당되어 있습니다.  
Random\_Seed에 의해서 동일 알고리즘의 성능차이가 발생할 수 있음에 유의바랍니다.

[문제1][0.6점] 태양광 발전량 예측 <https://www.kaggle.com/t/c94e73cb585b400fbb28770f4984f452>

최근 태양전지판을 설치하여 전기를 만들어 사용하는 가정집이 늘어나고 있다. 이는 초기 설비만 갓출 경우 반영구적으로 전기를 생산할 수 있고 전기생산과정에서 독성물질이 배출되지 않아 신에너지원으로 각광받고 있기도 하다. 그러나 태양광 에너지는 날씨의 영향을 많이 받아 생산량이 일정하지 않다는 단점이 존재한다. 최근 형준이네 집은 이와 같은 트렌드에 맞춰 태양전지판을 설치하였으나 내일의 전기 생산량을 예상할 수 없다는 어려움을 겪고 있다. 여러분이 지금부터 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 정보를 통해 태양광 발전량을 예측하는 인공지능 SW를 작성해주길 바란다.

학습 데이터로는 (기상 데이터 정보)와 해당 기상 상태에 따른 (태양광 발전량)을 함께 제공합니다.

Hour	Minute	DHI	DNI	WS	RH	T	TARGET
0	0	0	0	0	1.4	57.76	0
1	15	0	222	18	1.3	26.82	13
2	4	0	0	0	2.6	84.94	-6
3	18	30	0	0	1.5	89.18	2
4	1	30	0	0	2.4	84.35	12

테스트 데이터로는 (기상 데이터 정보)만 제공하며, 예측된 (태양광 발전량)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하셔야 합니다.

Hour	Minute	DHI	DNI	WS	RH	T
0	6	0	54	548	1.2	72.29
1	14	0	297	27	8.0	19.07
2	1	30	0	0	2.9	66.29
3	19	0	0	0	1.8	63.38
4	4	30	0	0	1.7	55.61

제공되는 (기상 데이터 정보)는 순서대로 Hour(시간), Minute(분), DHI(수평면 산란일사량), DNI(직달일사량), WS(풍속), RH(상대습도), T(기온), Target(태양광 발전량)입니다.

자, 그럼 테스트 데이터로 주어진 (기상 데이터 정보)를 가지고 (태양광 발전량)을 예측하여 봅시다.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받습니다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 0.1점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.  
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.  
(미 제출시 0.1점 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑤ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[문제2][0.6점] 교통량 예측 <https://www.kaggle.com/t/16c1eba7f94b418188e5db62df021498>

(코로나 이전) 특별자치도 제주에는 많은 관광객들의 사랑을 받는 곳이다. 많은 이들이 찾는 곳인 만큼 날씨와 유동인구 정보를 통한 교통량 예측을 통해, 대중교통의 차량과 배차간격 등을 조절하는 것은 매우 중요하다. 여러분이 지금부터 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 날씨와 유동인구 정보를 통해 교통량을 예측하는 인공지능 SW를 작성해주길 바란다.

학습 데이터로는 (날씨, 유동인구 정보)와 해당 정보에 따른 (교통량)을 함께 제공합니다.

	시도명	읍면동명	거주인구	근무인구	방문인구	총 유동인구	평균 속도	평균 소요 시간	평균 기온	일강수량	평균 풍속	월	일	년	교통량
0	0	26	76018.965	5009.811	36887.341	117916.117	40.467	21.667	17.750	0.000	2.725	9	25	2019	0.000
1	1	24	1232416.968	76043.955	451558.268	1760019.191	38.126	31.632	13.300	14.500	4.500	3	4	2018	1532.989
2	1	23	134260.946	8412.761	78177.981	220851.688	51.107	45.643	5.250	0.500	4.950	4	7	2018	710.107
3	1	25	289327.429	39102.424	287029.303	615459.156	33.067	48.000	26.300	0.500	1.500	8	14	2019	374.200
4	1	7	1116617.660	96560.651	524905.301	1738083.612	40.392	26.961	8.400	44.000	4.300	12	1	2019	535.882

테스트 데이터로는 (날씨, 유동인구 정보)만 제공하며, 예측된 (교통량)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하여야 합니다.

	시도명	읍면동명	거주인구	근무인구	방문인구	총 유동인구	평균 속도	평균 소요 시간	평균 기온	일강수량	평균 풍속	월	일	년
0	0	28	45737.828	8703.034	91775.025	146215.887	57.286	35.000	6.500	0.0	2.9	1	12	2020
1	1	21	289284.270	48392.749	179649.272	517326.291	50.500	58.912	18.833	0.0	2.3	6	10	2019
2	1	4	72684.218	45632.065	120931.098	239247.380	34.727	60.000	13.200	0.0	1.8	5	10	2019
3	1	12	235752.513	34212.805	186910.679	456875.997	37.167	26.500	11.700	0.0	4.8	3	19	2020
4	1	17	544250.956	64188.293	205182.677	813621.925	31.462	37.769	22.233	0.0	2.3	9	5	2018

제공되는 (날씨, 유동인구 정보)는 순서대로 시도명, 읍면동명, 거주인구, 근무인구, 방문인구, 총 유동인구, 평균 속도, 평균 소요 시간, 평균 기온, 일강수량, 평균 풍속, 월, 일, 년입니다.

자, 그럼 테스트 데이터로 주어진 (날씨, 유동인구 정보)를 가지고 (교통량)을 예측하여 봅시다.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받습니다.

>> csv파일에 한글이 포함되어 있습니다. 캐글 노트북에서 encoding 정보를 부여하여 파일을 읽기 바랍니다.

```
train = pd.read_csv("train.csv", engine='python',encoding='CP949')
```

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 0.1점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.  
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.  
(미 제출시 0.1점 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑤ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[문제3][0.6점] 통신사 고객 이탈 예측 <https://www.kaggle.com/t/c3503e255f114aac9bef4daedab0fd97>

통신 회사의 경우 신규 고객을 유치하는 동시에 수익 창출 기반을 늘리기 위해 계약 해지(=이탈)을 피하는 것이 중요합니다. 신규 고객이 이탈하는 이유를 살펴보면 더 나은 가격, 더 흥미로운 패키지, 불편한 서비스 경험 또는 고객의 개인적인 상황 변화와 같이 다양한 이유로 고객이 계약을 종료하게 됩니다. 고객 이탈 분석은 고객 이탈을 예측하고 이탈을 유발하는 근본적인 이유를 정의하는 기능을 제공합니다. 통신사는 기계 학습 모델을 적용하여 개별 고객을 기준으로 이탈을 예측하고 할인, 특별제안 또는 기타 만족을 주기 위한 대응 조치를 취하여 고객을 유지할 수 있습니다. 여러분은 지금부터 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 통신사 고객들의 이탈 여부를 예측하는 인공지능 SW를 작성해주길 바랍니다.

학습 데이터로는 (통신사 고객 정보)과 해당 고객들의 (최종 이탈 여부)를 함께 제공합니다.

Unnamed: 0	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
index																				
0	1869	7010-BRBUJ	Male	0	Yes	Yes	72	Yes	Yes	No	No internet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automatic)	24.10	1734.65	No
1	4528	9686-YGKVR	Female	0	No	No	44	Yes	No	Fiber optic	Yes	No	Yes	No	Month-to-month	Yes	Credit card (automatic)	88.15	3973.2	No
2	6344	8286-DQJDF	Female	1	Yes	No	38	Yes	Yes	Fiber optic	No	No	No	No	Month-to-month	Yes	Bank transfer (automatic)	74.95	2869.85	Yes
3	6739	6994-KERXL	Male	0	No	No	4	Yes	No	DSL	No	No	No	Yes	Month-to-month	Yes	Electronic check	55.90	238.5	No
4	432	2181-UAESM	Male	0	No	No	2	Yes	No	DSL	Yes	No	No	No	Month-to-month	No	Electronic check	53.45	119.5	No

테스트 데이터로는 (통신사 고객 정보)만 제공하며, 예측된 고객의 (이탈 여부)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하여야 합니다.

Unnamed: 0	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	...	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	
index																					
5	2215	4312-GVYNH	Female	0	Yes	No	70	No	No phone service	DSL	...	No	Yes	Yes	No	Yes	Two year	Yes	Bank transfer (automatic)	49.85	3370.20
10	3169	4578-PKJFZ	Male	0	Yes	Yes	52	Yes	No	DSL	...	Yes	Yes	Yes	Yes	No	One year	Yes	Electronic check	68.75	3482.85
13	1760	2511-MGRQY	Male	0	Yes	Yes	50	Yes	Yes	DSL	...	No	Yes	No	No	No	One year	No	Bank transfer (automatic)	54.90	2614.10
18	6508	8708-KPQHZ	Female	0	Yes	Yes	42	Yes	Yes	Fiber optic	...	No	No	No	Yes	Month-to-month	Yes	Electronic check	94.20	4186.30	
20	4693	0463-TXQAK	Male	0	No	Yes	52	Yes	Yes	No	...	No internet service	No internet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automatic)	25.60	1334.50
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	

제공되는 (통신사 고객 정보)는 순서대로 customerID, gender, SeniorCitizen[고령자], Partner, Dependents[부양가족], tenure[계약유지기간], PhoneService, MultipleLines[여러가인가입], InternetService[인터넷공급자], OnlineSecurity[온라인보안], OnlineBackup, DeviceProtection, TechSupport[기술지원], StreamingTV[TV스트리밍], StreamingMovies[영화스트리밍], Contract[계약형태-Month to Month, One year, Two year], PaperlessBilling, PaymentMethod, MonthlyCharges[월청구액], TotalCharges[전체기간청구액], Churn[이탈여부]입니다.

자, 그럼 테스트 데이터로 주어진 (통신사 고객 정보)를 가지고 고객의 (이탈 여부)를 예측하여 봅시다.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받는다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 0.1점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.  
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.  
(미 제출시 0.1점 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑤ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[문제4][0.6점] <https://www.kaggle.com/t/5ee6b736adce49019662dd9247d0efce>

영희는 보스턴에서 부동산중개업자로 일하기 시작한 초보 중개업자이다. 종종 집을 팔기 희망하는 고객이 적절한 판매가를 질문하곤 하지만 초보 중개업자인 관계로 이를 대답하기 쉽지 않을 때가 많다. 마침 영희 남편 철수는 IT회사 데이터 분석가로 일하고 있는 터라, 영희는 남편 철수에게 적절한 주택가격을 알려주는 프로그램을 만들어 달라고 부탁을 하려 한다. 여러분 역시 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 보스턴 주택가격을 예측해주는 인공지능 SW를 작성해 주기 바란다.

학습 데이터는 (주택 정보)와 (주택 가격)을 함께 제공합니다.

	CRIM	INDUS	NOX	RM	LSTAT	B	PTRATIO	ZN	CHAS	AGE	RAD	DIS	TAX	Price
index														
0	0.14150	0.0	6.91	0	0.448	6.169	6.6	5.7209	3	233	17.9	383.37	5.81	25.3
1	0.15445	25.0	5.13	0	0.453	6.145	29.2	7.8148	8	284	19.7	390.68	6.86	23.3
2	16.81180	0.0	18.10	0	0.700	5.277	98.1	1.4261	24	666	20.2	396.90	30.81	7.2
3	0.05646	0.0	12.83	0	0.437	6.232	53.7	5.0141	5	398	18.7	386.40	12.34	21.2
4	8.79212	0.0	18.10	0	0.584	5.565	70.6	2.0635	24	666	20.2	3.65	17.16	11.7

테스트 데이터는 (주택 정보)만 제공하며, 예측된 (주택 가격)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하여야 합니다.

	CRIM	INDUS	NOX	RM	LSTAT	B	PTRATIO	ZN	CHAS	AGE	RAD	DIS	TAX
index													
0	0.04932	33.0	2.18	0	0.472	6.849	70.3	3.1827	7	222	18.4	396.90	7.53
1	0.02543	55.0	3.78	0	0.484	6.696	56.4	5.7321	5	370	17.6	396.90	7.18
2	0.22927	0.0	6.91	0	0.448	6.030	85.5	5.6894	3	233	17.9	392.74	18.80
3	0.05789	12.5	6.07	0	0.409	5.878	21.4	6.4980	4	345	18.9	396.21	8.10
4	3.67822	0.0	18.10	0	0.770	5.362	96.2	2.1036	24	666	20.2	380.79	10.19

데이터로 제공되는 (주택 정보)는 순차적으로 자치시별 1인당 범죄율(CRIM), 비소매상업지역이 점유하고 있는 토지의 비율(INDUS), 0ppm당 농출 일산화질소 농도(NOX), 주택 1가구당 평균 방의 개수(RM), 인구 중 하위 계층 비율(LSTAT), 자치시별 흑인의 비율(B), 자치시별 학생/교사 비율(PTRATIO), 25000 평방 피트를 초과하는 거주지역의 비율(ZN), 찰수강 근접 여부(1-경계,0-아닌 경우)(CHAS), 1940년 이전에 건축된 주택 비율(AGE), 방사형 고속도로까지의 접근성 지수(RAD), 5개의 보스턴 직업센터까지의 접근성 지수(DIS), 1000달러 당 재산 세율(TAX) 이다.

자, 그럼 테스트 데이터로 주어진 (주택 정보)에 맞는 (주택 가격)을 예측하여 보자.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받는다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 0.1점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.  
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.  
(미 제출시 0.1점 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑤ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[문제5][0.6점] <https://www.kaggle.com/t/2aa002909aca4a28a64ad64c49707ee8>

최근 N사는 빅데이터분야와 데이터사이언스분야를 활용한 서비스가 많아지면서 해당 분야 개발자가 많이 필요해졌다. 이에 외부에서 직원을 추가 채용하는 부분을 고려하기 전 사내 부서 이동 제도를 통해 빅데이터분야와 데이터사이언스분야로 옮기고 싶어하는 직원들의 신청을 받기로 했다. 그러나 HR(인사팀)에서 근무하는 솔잎 양은 해당 직원들의 정보가 인공지능관련 부서로의 이동뿐만 아니라, 현 부서의 만족도가 낮아(분야, 임금, 동료, 기타 등등) 이직을 고려중인 직원으로 분류 가능하다는 분석결과를 도출하고 사내 직원들 중 현재 직장을 그만두고 새로운 일자리를 알아보는 직원을 예측하는 소프트웨어를 만들어 보려한다. 여러분 역시 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 이직하기 희망하는 직원을 예측해주는 인공지능 SW를 작성해 주기 바란다.

학습 데이터는 (직원의 개인 정보)와 직원의 (이직 희망 여부)를 제공합니다.

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_size	company_type	last_new_job	training_hours	target
index														
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	>20	NaN	NaN	1	36	1.0
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM	15	50-99	Pvt Ltd	>4	47	0.0
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Business Degree	<1	NaN	Pvt Ltd	never	52	1.0
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	STEM	>20	50-99	Funded Startup	4	8	0.0
5	21651	city_176	0.764	NaN	Has relevent experience	Part time course	Graduate	STEM	11	NaN	NaN	1	24	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

테스트 데이터는 (직원의 개인 정보)만 제공하며, 예측된 직원의(이직 희망 여부)는 submit.csv 파일로 저장하여 캐글 리더보드에 제출하여야 합니다.

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_size	company_type	last_new_job	training_hours	
index														
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	STEM	5	NaN	NaN	never	83	
9	699	city_103	0.920	NaN	Has relevent experience	no_enrollment	Graduate	STEM	17	10000+	Pvt Ltd	>4	123	
10	29452	city_21	0.624	NaN	No relevent experience	Full time course	High School	NaN	2	NaN	NaN	never	32	
11	23853	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	5	5000-9999	Pvt Ltd	1	108	
15	6588	city_114	0.926	Male	Has relevent experience	no_enrollment	Graduate	STEM	16	10/49	Pvt Ltd	>4	18	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

제공되는 (직원의 개인 정보)는 순서대로 직원의 고유 ID, 도시 코드, 도시 개발 지수, 직원의 성별, 직원의 데이터사이언스분야 관련 경험, 현 대학등록여부(풀타임, 파트타임, 없음), 학위(고졸, 대졸 등), 직원의 경력, 현 회사의 직원수, 고용 유형, 이전 회사 입사 년도, 이수한 교육 시간, 이직 희망 여부 이다.

자, 그럼 테스트 데이터로 주어진 (직원의 개인 정보)에 맞는 직원의 (이직 희망 여부)를 예측하여 보자.  
>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받는다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 0.1점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.  
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.  
(미 제출시 0.1점 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑤ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.