

제1회 외국인행정 공공데이터 활용 경진대회 제서			
공모분야	<b>「도전.공공데이터」아이디어 기획</b>		
과제명	<b>머신러닝 분류모델을 활용한 무사증 입국자 위험선별과 계절근로자 규모 예측 및 시각화</b>		
제안일자	2022.06.28.(예정)	대표자 성명	강나루
활용데이터 정보 ※복수기재가능 ※ 즐추가 가능	법무부 홈페이지, ICRM	법무부	외국인 출입국자현황 및 K-ETA 자료, 출입국 자료 등
	통계청 홈페이지(KOSIS)	통계청	농어업, 인구 등 각종 통계
	고용노동통계 홈페이지	고용노동부	농어업 종사자 통계자료
적절성 (공공데이터 활용 25점)	○공공데이터와 법무부의 누적된 외국인 데이터를 독립변수로 활용하여 머신러닝 및 통계 시각화로 무사증 외국인 위험선별 및 외국인 계절근로자 정책 수립 지원 기대.		
사회 문제 해결 방안 (문제해결가능성 20점)	○공공데이터 및 법무부 외국인 데이터를 활용하여 다양한 머신러닝 모델을 구축해보고 무사증 입국자의 범죄, 자격외활동 및 체류기간도과 등을 입국 전 위험선별 하고자 함.  ○농어촌 인력난을 해결하기 위해 계절근로자 규모의 추정치 계산 및 이를 지자체별 시각화 해보고, 계절근로자 배정심사협의회에 제공하여 참고 자료로 활용하는 등 공공데이터를 활용해 농어촌 인력 수급에 도움이 되고자 함.		
차별화 전략 (독창성 20점)	○법무부의 풍부한 외국인 데이터를 타 부처의 공공데이터와 결합해 사회 문제에 대응하고, 머신러닝으로 새로운 무사증 외국인 선별 기준 수립의 근거 자료 제공.  ○농어촌 인력난을 겪고 있는 지자체에게 통계 분석을 통해 계절근로자 도입 규모 근거를 다양화, 객관화하여 제공할 수 있음.  ○국민이 체감할 수 있는 농어촌 인력난, 농수산물 가격 변동, 불법체류자 증가 문제에 공감.		
아이디어 구체성 (실현가능성 15점)	○공공데이터, K-ETA, ICRM, 해외 자료를 활용한 머신러닝 학습 선별로 무사증 입국자 중 입국목적이 불분명한 외국인을 위험선별(스크리닝)하고자 함.  ○향후 농어업 계절근로자 수를 다중회귀분석으로 추정해보고, 계절근로자 제도의 규모를 데이터 시각화로 제시하여 계절근로자 배정심사협의회에서 참고 자료로 활용 가능.		
기대성 (발전가능성 20점)	○머신러닝을 활용한 입국자 위험선별로 불법체류율 감소 및 데이터 기반의 신뢰도 높은 출입국 행정 실현 기대.  ○지자체별 계절근로자 규모 시각화를 활용하여 계절근로자의 적재적소 공급으로 농어촌 인력난 해소와 신생 농어업법인 설립 기대, 농수산물 가격 변동 축소 기대.  ○장기적으로 무사증 입국자를 선별해 비허가 자격외활동을 감소시키고, 계절근로자 모델을 제조업 등 타 분야에 확대 적용하여 합법체류자의 적절한 인력수급 예측 기대.		

## 첨부1

# 머신러닝 학습의 무사증 입국자 위험선별 운영 절차

### 절차

### 주요사항

### 부가사항

모델개발

머신러닝 모델 선정 및 데이터 선별

- 공공데이터 데이터 선별
- 데이터 전처리 학습 필요



학습  
데이터  
구축

**무사증 입국자 학습데이터 구축**  
(공공데이터, 법무부 데이터, 해외 현지  
데이터 활용)

- K-ETA, 출입국기록 등  
법무부 외국인 데이터와  
공공데이터 활용  
(EPS-TOPIK 통계 등)
- 태국 시도별 실업률 등  
현지 통계자료 반영
- 태국 주민등록번호를  
활용한 지역별 분류 예정



분석 모델  
구축

**엔터티 임베딩**  
임베딩 전처리로  
데이터 변환  
(변수 간 관계를  
N차원 변수로 표현)

**무사증 입국자  
선별 모델 선정**

- 머신러닝 여러 모델을 비교,  
융합하여 최적 모델 선정



모델 평가

머신러닝 모델 간  
비교 및 최적화

학습데이터와  
검증데이터로  
분류하여 **모델 평가**

- Random Forest 등 위험선별  
적중률 비교 평가



시스템  
구축

**무사증 입국자 위험선별**  
**K-ETA 및 입국심사, 체류 관리 활용**

- 무사증 입국자를 학습데이터로  
자동 위험 선별(% 표출)
- 실시간 유입 데이터 선별



자동  
리모델링

**기존 학습데이터**  
실시간 유입  
위험선별 활용

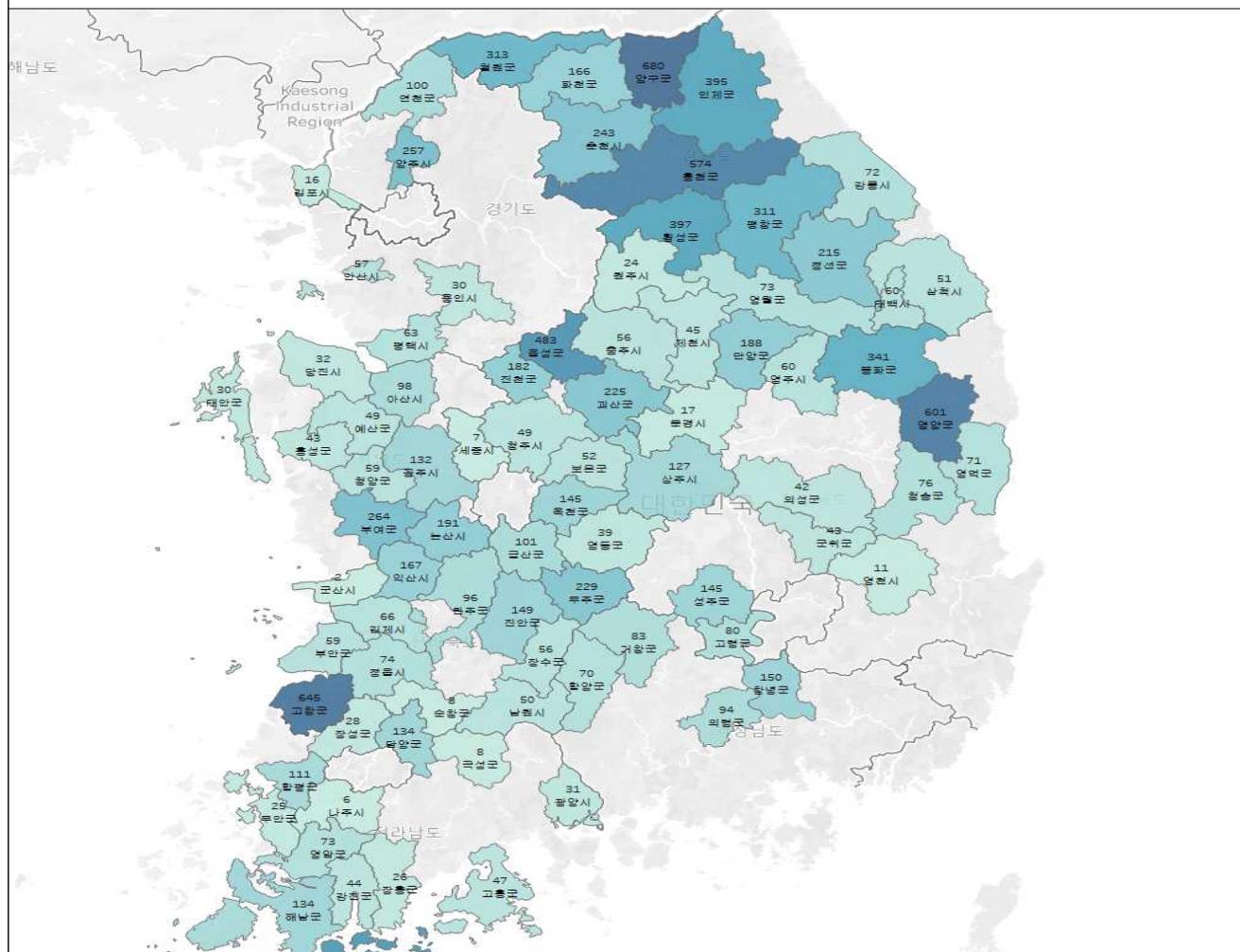
**시스템 구축  
이후 데이터**  
자동 리모델링  
되어 학습 검증

- 무사증 입국자 위험선별  
플랫폼 활용 및 자동  
리모델링 학습 및 검증
- 불법체류 감소 및 데이터  
기반 출입국 행정 실현으로  
신뢰도 향상 기대

첨부2

## 계절근로자 모델링 방안-데이터시각화, 국내시도별로

## 2022년도 지역별 계절근로자 배정 인원 데이터 시각화 자료



## 계절근로자 수 다중회귀분석(예시)

종속변수(Y축)	독립변수(X축)	다중 선형 회귀분석 방정식	해석 및 활용방안
계절근로자 수(A)	고용허가제(E-9) 농림어업 쿼터(B) 농가구입가격지수 2)노무비(C)	A=1.75688B+144.88 33C-30221.6	고용허가제 농림어업 쿼터와 농가구입 가격지수 중 노무비는 계절근로자 수에 유의한 영향을 미치며 법무부 데이터1), 지자체 데이터를 반영하여 계절근로자 규모를 추정하고 시각화할 수 있음.

#### a. 졸속변수·계절군로자수

모형	비표준화 계수		표준화 계수 베타	t	유의확률	공선성 통계량		3)
	B	표준화 오류				공차	VIF	
1	(상수)	-30221.578	2814.272		-10.739	<.001		
	E-9 농림어업	1.757	.423	.505	4.152	.009	.446	2.241
	가격지수	144.883	32.216	.547	4.497	.006	.446	2.241

1) 불법체류자 증감, 다문화 농가인구 등 데이터를 지역별 통계에 반영하여 종속변수(계절근로자 수) 예측  
2) 노연경계연령의 간접 및 경연활동에 틀이된 노동력이 2015년 기준 가령진술 출처: KOSIS

3) R<sup>2</sup>(.083), R<sup>2</sup>(.067), Adj-R<sup>2</sup>(.528), F(73.107), 유의 학률(.001), Durbin-Watson(1.968)