

지역축제 방문객 예측 모델

소프트웨어융합학과

2017103728

배이지

과제 목표 수정

- 독립변수: 날씨, 축제기간, KTX 역 유무, 요금정보, 인스타그램 언급횟수, 긍정 리뷰 수, 부정 리뷰 수
- 종속변수: 방문객 수, 내국인 외국인 비율, 경제효과, 축제등급
- 결과
 - 1) 세 가지 모델 다중선형회귀, 인공신경망, 서포트벡터 머신을 이용해 예측 후 회귀식, R-Sqaure, 예측정확도를 비교해 가장 성능이 뛰어난 모델을 선정한다.
 - 2) 웹페이지
 - 선택페이지: 축제 명, 날짜, 날씨 및 기간 설정
 - 결과페이지: 예측 방문객수, 내외국인 비율, 경제효과, 축제등급을 보여준다.
또한 인스타그램 언급 및 감정분석 결과를 나타낸다.

대분류	중분류	소분류	설명	날짜
표본선정	사용할 데이터 정리			○
데이터 수집 및 가공	축제 데이터 결측값 확인 및 채우기	날짜 및 시기, 등급, 세부지역 및 축제이름 통일	Python	○
	기상청 데이터 연결하기			4/10
	입장료 유무			4/10
	KTX 노선 연결하기			4/11
	인스타그램 감정분석	인스타그램 언급횟수, content 크롤링	Python	4/13
		Content 감정분석	감정사전	4/18
변수 정의	독립 변수 및 종속변수 확정하기			4/18
데이터 탐색	기초 통계 및 변수 간 상관관계 계산	인스타그램 언급횟수, 긍정리뷰수, 부정리뷰수		4/19
모델구축 및 평가	데이터 분할			5/18
	변수조합탐색		R stepwise	5/18
	예측모델훈련	데이터 적합도 그래프		5/18
		다중선형회귀, 인공신경망, 서포트벡터머신	R	5/18
	예측모델평가	R-square		5/18
		예측 정확도		5/18
웹페이지구축	선택 페이지	축제명, 날씨, 기간, 날짜 등을 선택	장고	6/15
	결과 페이지	예측 방문객 수, 내외국인 비율, 경제효과, 축제 등급		6/15
		이전 년도 방문객 수 및 정보		6/15
		인스타그램 분석 결과		6/15