

2021 데이터마이닝

유기동물 생존 확률 예측

인공지능융합전공 2019311195 김지유
인공지능융합전공 2019311492 이나은

목차

1

주제소개

2

데이터소개

3

모델소개

4

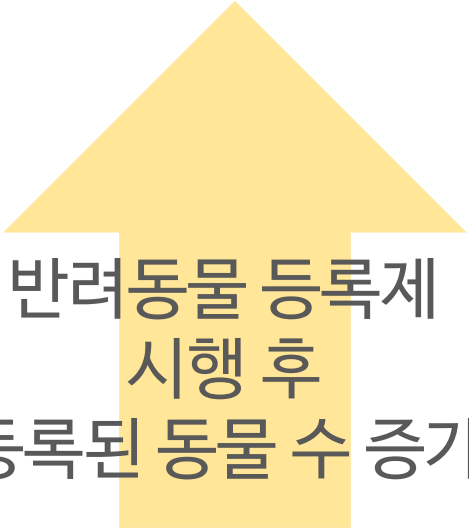
결과 및 시연

5

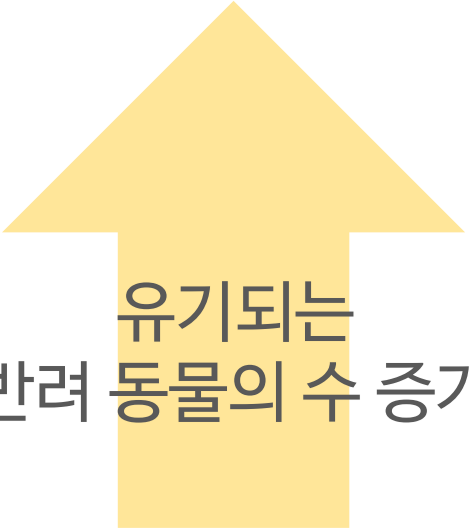
시사점 및 의의

1

주제소개



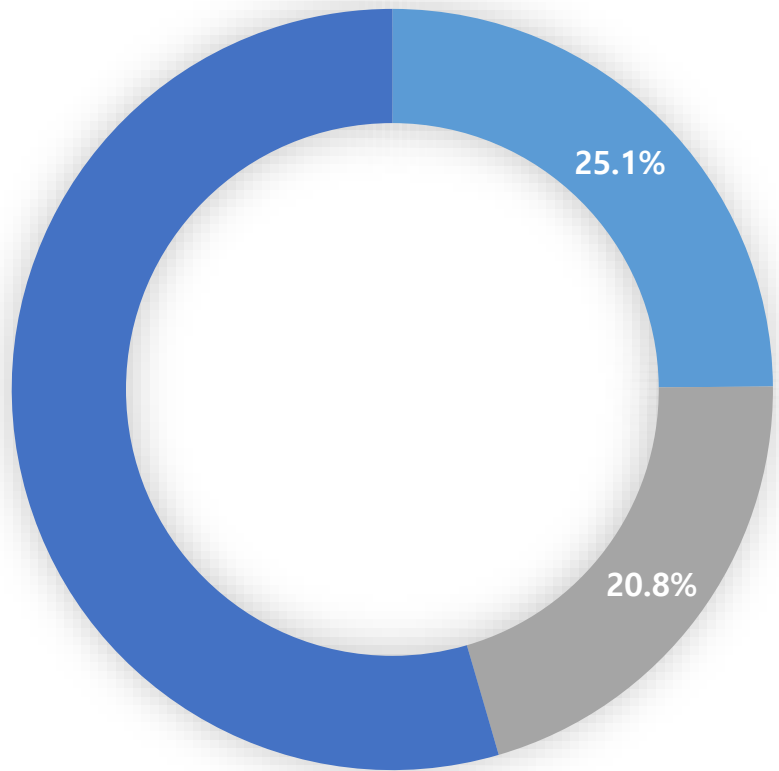
반려동물 등록제
시행 후
등록된 동물 수 증가



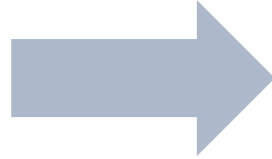
유기되는
반려 동물의 수 증가

1

주제소개



■ 자연사 ■ 안락사



- 반려 동물 유기로 인한 자원 소모
- 생명체가 강제로 죽음 맞이

1

주제소개

- ☐ 능동적인 해결책

반려견의 유실과 유기를 막는 것

- ☒ 수동적인 해결책

이미 유실 또는 유기된 동물의 입양률 높이기

1

주제소개

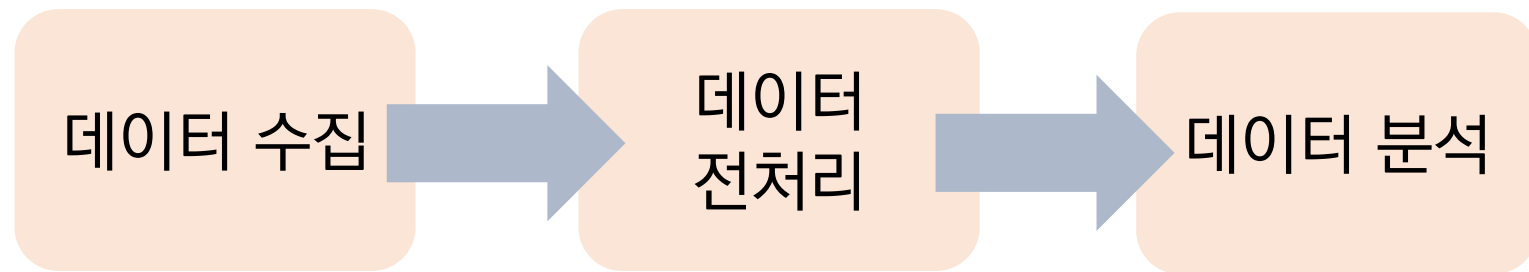


“유기견의 생존 확률 예측 프로그램 제작”

생존 확률이 낮은 동물 순으로 입양 공고를 띄우는 웹페이지 제작

2

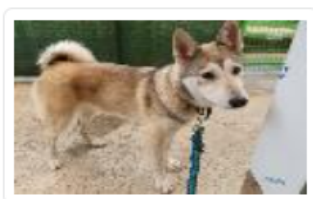
데이터 소개



2

데이터 수집

동물보호관리시스템 유실/유기 동물 공고 페이지 크롤링



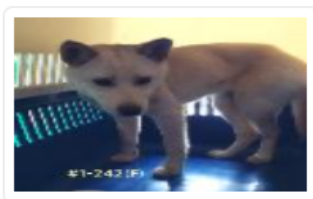
자세히 보기

공고번호 경남-창원1-2021-00249
접수일자 2021-05-20
품종 라이카
성별 암컷
발견장소 의창구 북면 마산리..
특징 K-5-1-(3)..
상태 종료(반환)
공고기간 2021-05-20 ~ 2021-05-29



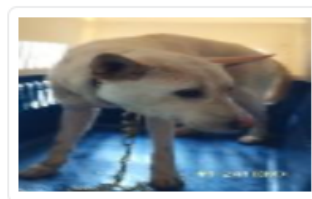
자세히 보기

공고번호 경남-고성-2021-00242
접수일자 2021-05-20
품종 푸들
성별 수컷
발견장소 대가면 금산리
특징 1-243번
상태 공고중
공고기간 2021-05-20 ~ 2021-05-27



자세히 보기

공고번호 경남-고성-2021-00241
접수일자 2021-05-20
품종 믹스견
성별 암컷
발견장소 대가면 금산리
특징 1-242번
상태 공고중
공고기간 2021-05-20 ~ 2021-05-27



자세히 보기

공고번호 경남-고성-2021-00240
접수일자 2021-05-20
품종 믹스견
성별 수컷
발견장소 거류면 당동리 59..
특징 1-241번
상태 공고중
공고기간 2021-05-20 ~ 2021-05-27

2

데이터 수집

1. 동물의 정보	
광고번호	경남-창원1-2021-00249
축종	개
품종	라이카
털색	회색+흰색
성별	암컷
중성화 여부	아니오
특징	K-5-1-(3) 회색+흰색, 까만코, 파란색 목줄착용.



20,132개의 row와 9개의 column으로 구성된 dataset

2

데이터 전처리

상태	공고번호	축	품종	털색	성	중성화 여부	특징	접수일시
종료(자연사)	경기-광주	고양이	한국 고양이	검백색	수컷	아니오	5-42, 3주령	2020-05-09
종료(입양)	대구-달성	고양이	한국 고양이	흰색	수컷	아니오	1주령 추정	2020-05-09
종료(반환)	부산-사하	개	말티즈	흰색	수컷	미상	흰색	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	흑백	수컷	아니오	어린 고양이	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	삼색	암컷	아니오	어린 고양이	2020-05-09
종료(자연사)	울산-남구	고양이	한국 고양이	흰색 검정	암컷	아니오	인공포유	2020-05-09
종료(안락사)	제주-제주	개	믹스견	흰색	수컷	미상	(개체관리됨)	2020-05-08
종료(입양)	경남-김해	개	믹스견	갈흠	수컷	미상	630, 6월추	2020-05-08
종료(입양)	경남-사천	개	웰시 코기	황색 바탕	암컷	아니오	관리 잘 됨	2020-05-08

1. **상태** : 기증, 방사와 같이 생존 또는 사망으로 분류하기 애매한 값을 가진 행 삭제
2. **공고번호** : 행정시 또는 자치시만 남김 (ex. 경기, 대구 등)
3. **축** : '개'를 제외한 나머지 행 삭제

2

데이터 전처리

상태	공고번호	축	품종	털색	성	중성화 여부	특징	접수일시
종료(자연사)	경기-광주	고양이	한국 고양이	검백색	수컷	아니오	5-42, 3주령	2020-05-09
종료(입양)	대구-달성	고양이	한국 고양이	흰색	수컷	아니오	1주령 추정	2020-05-09
종료(반환)	부산-사하	개	말티즈	흰색	수컷	미상	흰색	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	흑백	수컷	아니오	어린 고양이	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	삼색	암컷	아니오	어린 고양이	2020-05-09
종료(자연사)	울산-남구	고양이	한국 고양이	흰색 검정	암컷	아니오	인공포유	2020-05-09
종료(안락사)	제주-제주	개	믹스견	흰색	수컷	미상	(개체관리됨)	2020-05-08
종료(입양)	경남-김해	개	믹스견	갈흠	수컷	미상	630, 6월추	2020-05-08
종료(입양)	경남-사천	개	웰시 코기	황색 바탕	암컷	아니오	관리 잘 됨	2020-05-08

4. **품종** : 동물보호시스템에서 제공하는 품종 리스트에 포함된 행만 남김
5. **털색** : 자주 등장하는 대표적인 색들이 포함된 행만 남김
6. **성** : '미상'인 값을 가진 행 삭제

2

데이터 전처리

상태	광고번호	축	품종	털색	성	중성화 여부	특징	접수일시
종료(자연)	경기-광주	고양이	한국 고양이	검백색	수컷	아니오	5-42, 3주령	2020-05-09
종료(입양)	대구-달성	고양이	한국 고양이	흰색	수컷	아니오	1주령 추정	2020-05-09
종료(반환)	부산-사하	개	말티즈	흰색	수컷	미상	흰색	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	흑백	수컷	아니오	어린 고양이	2020-05-09
종료(입양)	경기-남양주	고양이	한국 고양이	삼색	암컷	아니오	어린 고양이	2020-05-09
종료(자연)	울산-남구	고양이	한국 고양이	흰색 검정	암컷	아니오	인공포유	2020-05-09
종료(안락)	제주-제주	개	믹스견	흰색	수컷	미상	(개체관리됨)	2020-05-08
종료(입양)	경남-김해	개	믹스견	갈흠	수컷	미상	630, 6월추	2020-05-08
종료(입양)	경남-사천	개	웰시 코기	황색 바탕	암컷	아니오	관리 잘 됨	2020-05-08

7. **특징**: 긍정/부정 단어 사전 구축 후 부정 점수, 긍정 점수로 변환

⇒ 최종적으로 object 데이터 타입을 더미 변수로 변환

⇒ 8,359개의 row와 86개의 column을 갖는 csv 파일 생성

2

데이터 분석

1. 보호소의 위치에 따른 입양률 분석

- 높은 입양률 : 서울, 대전, 광주 => 평균 80%
- 낮은 입양률 : 울산, 세종, 제주 => 평균 25%

2. 중성화 여부에 따른 입양률

- 중성화 수술을 받은 개 : 0.85
- 중성화 수술을 받지 않은 개/미상 : 0.47

2

데이터 분석

3. 라벨(생존/사망)과 피처의 상관관계 분석

- 부정 점수와 라벨 : 음의 상관관계
- 긍정 점수와 라벨 : 양의 상관관계
- 중성화 수술을 받은 경우와 라벨 : 양의 상관관계

2

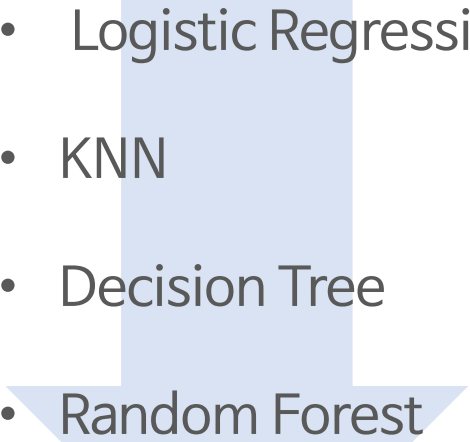
데이터 분석

4. 생존 유기 동물의 특징, 사망한 유기동물의 특징을 워드 클라우드로 분석



3

모델소개

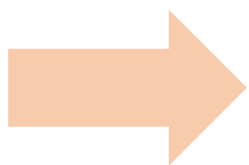
- 
- Logistic Regression
 - KNN
 - Decision Tree
 - Random Forest

최종 모델 : Random Forest

4

결과 및 시연

	precision	recall	f1-score	support
0	0.77	0.80	0.78	870
1	0.77	0.74	0.75	802
accuracy			0.77	1672
macro avg	0.77	0.77	0.77	1672
weighted avg	0.77	0.77	0.77	1672



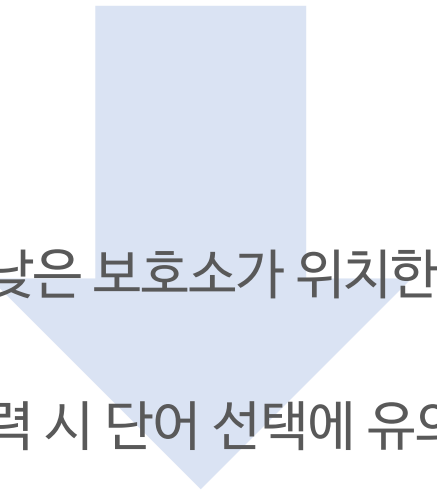
예측한 결과를 바탕으로 생존 가능성이 낮은 동물을 상단에 배치하는 웹 사이트 제작

4

결과 및 시연

데이터 분석 결과,

보호소가 위치한 **도시**, 공고자가 임의로 작성하는 **특징**, **중성화 여부**가 생존에 가장 큰 영향을 미침

- 
- ☑ 국가 차원 : 생존율이 낮은 보호소가 위치한 도시에 특별 조치
 - ☑ 보호소 차원 : 특징 입력 시 단어 선택에 유의, 중성화 수술 확대

5

시사점 및 의의

현재 유기동물의 생존 확률을 토대로 구축된 유기동물 공고 사이트는 없다.

따라서, 유기동물의 생존 확률 예측 프로그램을 이용하여 **생존 확률이 낮은 유기동물의 치료 비용에 대한 후원 공고**를 올리거나 그들을 **입양 공고의 상단에 자주 노출시키는 방법**을 이용하여 궁극적으로 유기동물의 생존 확률을 높일 수 있을 것으로 기대한다.

THANK YOU

감사합니다!