

2021 전주환 포트폴리오

이메일 : powwow94@hanmail.net

전화번호 : 010-2616-6435



<https://ninano.cf/>

과정명

제 2020-429 호

훈련과정 이수증

성명 : 전주환
주민등록번호 : 941001-1*****

훈련과정 : 빅데이터분석
훈련직종 : 빅데이터 분석
훈련기간 : 20.09.14 ~ 21.03.15

상기 교육 훈련생은 해당
평가를 통해 훈련과정을

합격 년월일 : 2021.03.15

하이미디어

□ 훈련과정 이수 결과

비MCS 교과

교과목명	주요구성내용 (단원명)	이수여부	비고
입학식(0P)	오리엔테이션	P	2h
웹서비스 개발	HTML5 (메이커에이스)	P	80h
	HTML(8h) / CSS(8h) / JavaScript(24h) (프론트엔드프로그래밍)	P	40h
	JAVA(80h) / JSP/Servlet(40h) / Spring(40h) (서버사이드프로그래밍)	P	160h
	웹서비스 구현 (웹 프로젝트)	P	80h
통계방법론	기초통계(8h) / 확률 및 분포(8h) / 추론통계(8h) / 회귀분석(16h) (통계방법론)	P	40h
정렬6개월형 데이터 분석	Excel(16h) / R(80h) / Python(80h)	P	176h
	PowerBI(16h) / R(16h) / Python(24h) (데이터시각화)	P	56h
머신러닝을 활용한 데이터 분석	TensorFlow(24h) / Keras(16h) (머신러닝을 활용한 데이터 분석)	P	40h
빅데이터 프로젝트	모바일웹(8h) / 데이터수집(8h) / 전처리(24h) / 데이터분석 모델링 (40h) / 프로젝트 시각화 리포 팅(40h) (오픈데이터분석프로젝트)	P	120h
	서비스구현 (웹어플리케이션)	P	160h
	그룹PR (프로젝트발표)	P	6h

* 이수여부는 100점 만점 중 60점 이상인 경우 Pass(P), 미만인 경우 Fail(F)

기업맞춤형 채용연계 국가기간전력산업직종훈련

4차 산업혁명 선도 인력 양성을 위한

**빅데이터를 활용한
측분석전문가**

데이터분석리브러리 | 머신러닝
현장연수 프로젝트 | 통계분석
데이터시각화 | 서버사이드개발
프론트엔드개발 | 하둡분산처리

하이미디어
컴퓨터교육원

**BIG
DATA**

기업맞춤형 국가기간전력
통계, 분석, 시각화, 머신러닝, 프로젝트

취업연계

수료 교육

* 빅데이터를 활용한 예측 분석 전문가 양성과정
2020.09.24 ~ 2021.03.15

- HTML/CSS/JS를 활용한 개인 홈페이지 제작
- Oracle Database를 활용한 데이터저장 관리
- R를 활용한 기초 통계분석
- PYTHON를 활용한 머신러닝/딥러닝/데이터시각화

INDEX

0 1 . [분 석 모 델] K a g g l e , D a c o n

0 2 . 공 모 전 세 종 시 주 택 시 장 특 성 분 석

0 3 . [팀 프 로젝 트] 손 감 지 및 미 디 어 컨 트 롤

Subject	Model	Metrics	Data set
Boston House Prices	Regression (Ridge / Lasso)	rmse: 4.923	506 rows , 13 cols
House Prices: Advanced Regression Techniques	Regression (Ridge, Lasso) ElasticNet	rmse: 1.0276 rmse: 1.0147	Dataset : 4 files, 163 cols
Bike Sharing Demand	Regression (Ridge / Lasso)	rmsle:0.354 rmse:50.222 rmsle:0.330 rmse:53.362 rmsle:0.319 rmse:47.215	10886 rows, 12 cols
Titanic: Machine Learning from Disaster	DecisionTreeClassifier RandomForestClassifier LogisticRegression	auc: 0.7877 auc: 0.8324 auc: 0.8659	891 rows, 12 cols
Credit Card Fraud Detection	DecisionTreeClassifier RandomForestClassifier	f1: 0.7626 f1 : 0.8930	85290 rows,5cols
Pima Indians Diabetes Database	RandomForestClassifier	f1: 0.6931	768 rows,9 cols
Digit Recognizer(MNIST)	CNN	loss : 0.0024 , acc: 0.9723	1797rows,64 cols
Breast Cancer Wisconsin (Diagnostic) Data Set	Classification XGBoost/ LightGBM	f1 : 0.9677 f1: 0.9615	357 rows, 212cols
Auto MPG	XGBoost	Mse: 0.0129	397 rows, 9 cols
Cat and Dog	CNN	epoch:100 , batch-size:32	image : 217.88 MB

운동 동작 분류 AI 경진대회
월간 데이터 11 | 헬스 데이터 | Logloss | 분류

상금 : 100만원
2021.01.11 ~ 2021.02.22
760팀 | D-2

운동 동작 분류 AI 경진대회
월간 데이터 11 | 헬스 데이터 | Logloss | 분류

상금 : 100만원
2021.01.11 ~ 2021.02.22 17:59
767팀 | 마감

대회안내 | 데이터 | 코드 공유 | 토론 | 리더보드 | 제출

Public | Private | Ranking Chart

01.11 대회 시작

개요 | 규칙

#	팀	팀 멤버	최종점수	제출수	등록일
67	highl_B		0.80015	17	3일 전
1	hahaha		0.46096	93	6일 전
2	ghghdld		0.46687	53	3일 전
3	Jamm		0.47722	39	13일 전
4	세아이부지		0.48059	6	16일 전
5	KPH		0.48673	75	3일 전
6	국경원 요원		0.49255	23	4일 전

적용기술

Language : Python

OpenAPI : Numpy, Pandas, Scikitlearn

Visualization : Matplotlib, seaborn

IDE : Google Colab, JupyterNotebook

주최기관 : **DACON**

대 회 명 : 운동 동작 분류 AI 경진대회

U R L :

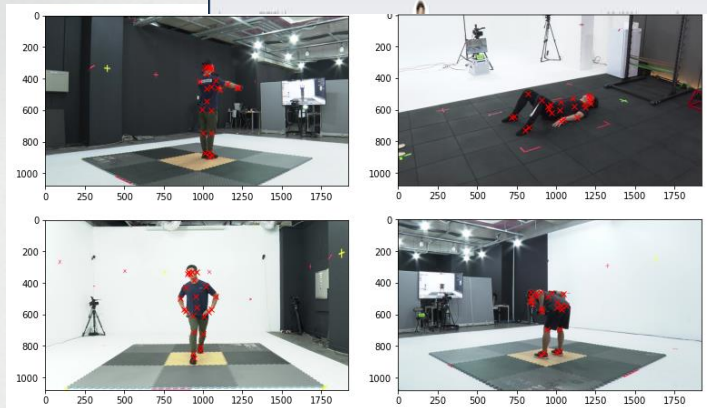
<https://dacon.io/competitions/official/235689/overview/>

분석과제 : 3축 가속도계(accelerometer)와 3축 자이로스 코프(gyroscope)를 활용해 측정된 센서 데이터에 머신러닝 알고리즘을 적용해 운동 동작 인식 알고리즘 개발

분석내용 : 1875000 건의 동작 데이터를 id별(3125건)로 묶어 합, 평균, 중앙값, 최소값, 최대값, 표준편차, 분산을 이용하여 데이터프레임을 재구성하여 분석함.

최종 선택 값 : lgbm, xgboost, randomforest model을 앙상블(soft voting) 하여 predict_proba 를 도출해 냄.

등수 : 67/839



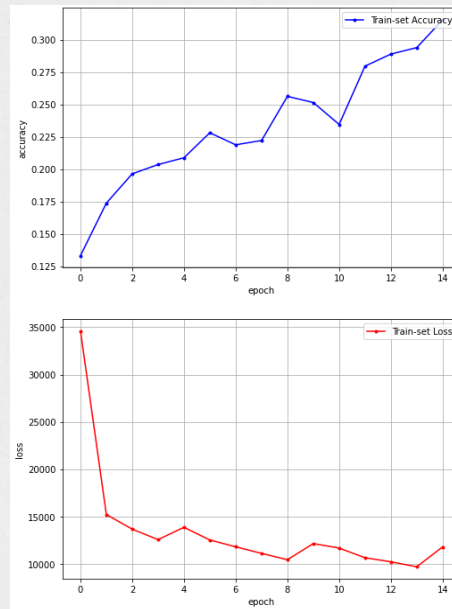
적용기술

Language : Python

OpenAPI : Keras, Tensorflow, Pandas, numpy, glob

Visualization : Matplotlib, tqdm

IDE : Google Colab, JupyterNotebook



주최 : 슬릭, 데이터연구소

주관 : DACON

대 회 명 : 모션 키폰트 검출 AI 경진대회

U R L :

<https://dacon.io/competitions/official/235689/overview/>

분석과제 : 머신러닝/딥러닝을 이용한 Motion keypoint detection 알고리즘 개발 및 스마트 헬스케어 산업에 적용 가능한 데이터 분석

분석내용 : CNN모델을 이용한 모션 키폰트 검출

등수 : 27/508

팀명

Ensemble

개요

- 세종시 주택 시장 특성을 쉽고 명확히 보이도록 시각화 모델 제시

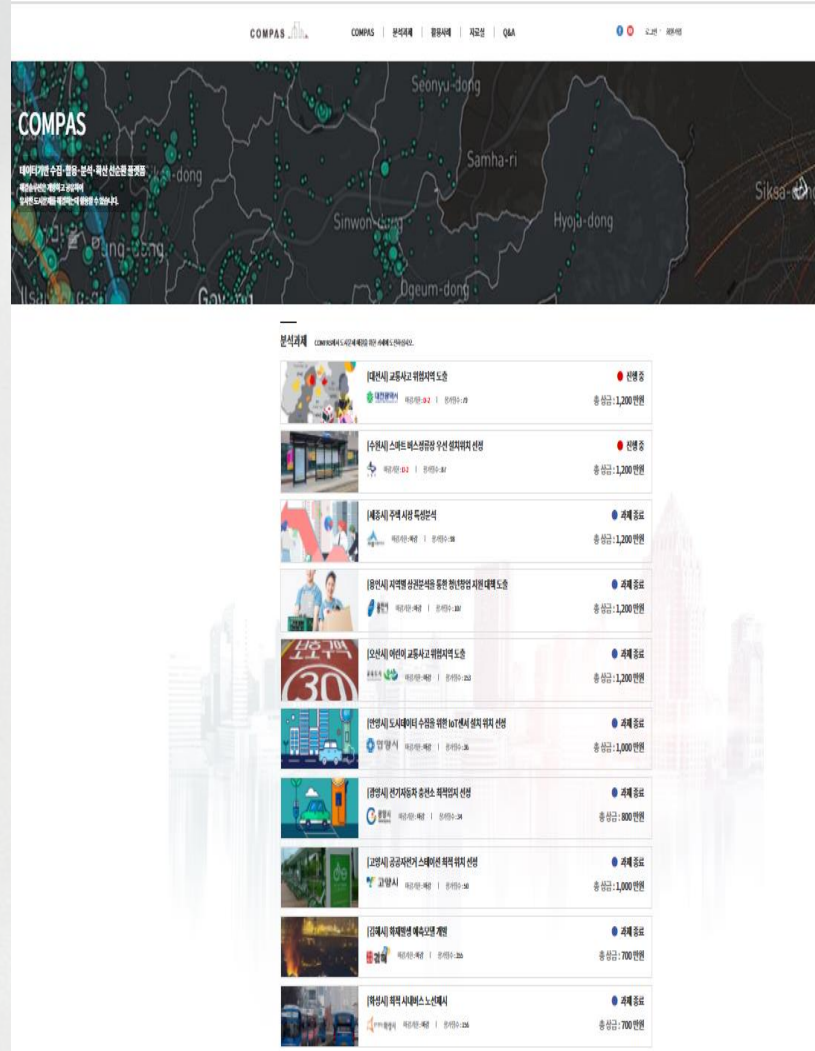
설명

- 세종시 주택 시장 특성을 직관적으로 이해 할 수 있는 다양한 모델을 도출

- 실거래가, 전입전출 등 거주인구 정보, 건물 등의 공간데이터를 활용 및 분석하여 그래프 및 지도를 시각화

적용기술

- Language : Python, R
- OpenAPI : Pandas, Numpy, geopy
- Visualization : plotnine, Matplot, seaborn, folium
- IDE : JupyterNotebook, Colab



Project Analysis

- 필요한 feature로 dataframe을 재구성하고 matplotlib, seaborn 을 통해 bar graph, pie chart 시각화

- folium 모듈을 사용하여 관공서, 교육시설 위치를 지도 위에 시각화

- One-hot Encoding하여 거래금액과의 상관관계를 보기 위한 heatmap을 생성

- 2017년, 2020년 전용면적 data set 재구성, R을 이용하여 data를 전용면적별로 구분

결론

: 주택 유형별로 형성되어 있는 매매/전월세 가격에 미치는 요인을 상관분석으로 도출한 결과, 대부분의 가구에서는 전용면적이 가격결정에 가장 큰 영향을 주는 요인이었으나, 일부지역에서는 소득 수준, 인프라, 교통과 인접성, 토지 가격인 것으로 확인되었다. 또한, 관공서, 초등학교, 중학교, 고등학교가 위치해 있는 부근을 중심으로 부동산 거래가 활발, 가격이 높게 나타났다.

세종시 주택 시장 특성분석 시각화

주택가격 결정요인 분석

팀명: Ensemble



5. 시각화



매매 실거래가 상관분석을 통하여 아파트 연립다세대, 단독 단가구 오피스텔 가격에 가장 큰 영향을 미치는 요인을 분석하고 주택 유형별 특성을 도출하였다.

실거래가 데이터를 encoding 하여 불러오고 object 타입으로 되어있던 거래금액을 int 형으로 변환 하였다. 가격에 영향을 미치지 않는 불필요한 컬럼을 삭제하고 One hot Encoding하여 Heatmap으로 상관분석하였다.

1) 아파트 (그림)

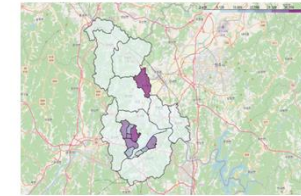
아파트 가격에 영향을 미친 요인은 상관성이 높은 순서대로 전용면적, 중 별 표층으로 나타났다. 지역은 조지원읍, 생원동 다정동이 아파트 가격과 상관성이 높은 것으로 나타났다.

2) 연립다세대 (그림)

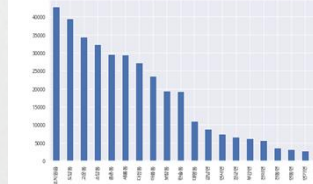
연립다세대 가격에 영향을 미친 요인은 상관성이 높은 순서대로 전용면적, 대지면적으로 나타났다. 지역은 반곡동, 조지원읍이 연립다세대 가격과 상관성이 높은 것으로 나타났다.



5. 시각화



2020년 자료를 바탕으로 지역별 인구 분포를 지도에 나타내었다. 조지원읍에 거주하는 인구가 가장 많게 나타났고 면 지역으로 갈수록 인구의 분포가 줄어드는 것을 알 수 있다.

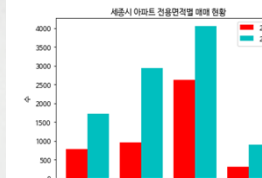


연령별 인구현황 데이터를 가지고 읍면동의 dataframe을 추출하였다. 위치 정보 geojson 형식 경계를 이용하여 위경도로 바꾸고 세종시 중심부의 위도 경도를 가지고 map을 출력하였다.

세종시 법정구역 세대원 분포도를 나타내기 위해 dataset을 재구성하고 index를 설정해 bar graph로 시각화하였다.



5. 시각화



상관분석 결과 아파트 거래금액에 가장 많은 영향을 미친 요인으로 전용면적과 지역으로 나타나 지역별 인구분포와 거래 면적을 토대로 왜 그러한 결과가 도출되었는지 살펴보았다.

그 결과, 2020년 기준 가장 많이 거래된 아파트 경수는 소형, 중형 경수이다. 조지원읍, 생원동, 도원동에는 1인, 2~3인 가구가 많이 거주하고 있으므로 소형, 중형경수의 아파트 거래가 많은 것으로 나타났다.



가장 많이 거래된 연립다세대의 평형은 조소형과 소형 경수이다.

가격과 연관이 있는 지역으로는 반곡동과 조지원읍이 있었는데 1인, 2~3인 가구가 많았던 조지원읍의 조소형, 소형 경수 비중이 높음에 비해 반곡동의 거래가 높았던 이유는 획지형 주택지 분양 등 다른 택지들에 비해 교통이 발달하여 인접성 이용이해 인구가 유입되고 있는 과정이고 조소형, 소형경수를 선호하는 수요가 많아져 거래건수가 많았던 것으로 나타났다.

2017년 (1~12월)과 2020년 (1~12월)의 전용면적 데이터 세를 재구성하고 데이터 처리를 위해 CSV로 데이터를 내보내고 내보낸 csv를 R을 이용하여 전용면적이 40㎡ 이하 일 때 조소형, 40 ~ 60㎡ 소형, 60 ~ 85㎡ 중형, 85㎡ 초과를 대형으로 구분하였다. 내보낸 csv파일은 인코딩 문제로 발람영을 직접 수정하였다. (전용면적(㎡) -> 전용면적)



프로젝트 명

WAVING

개요

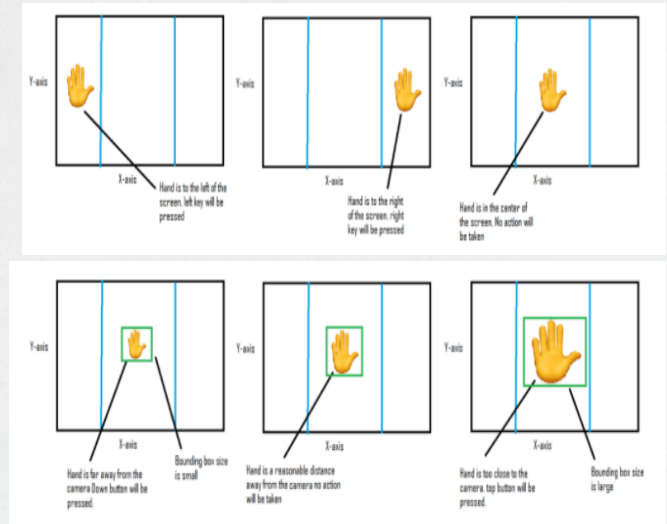
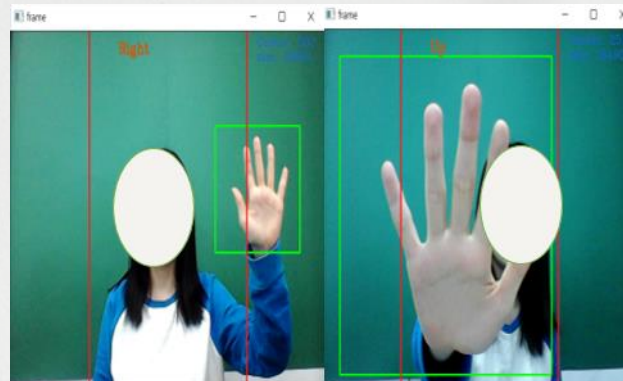
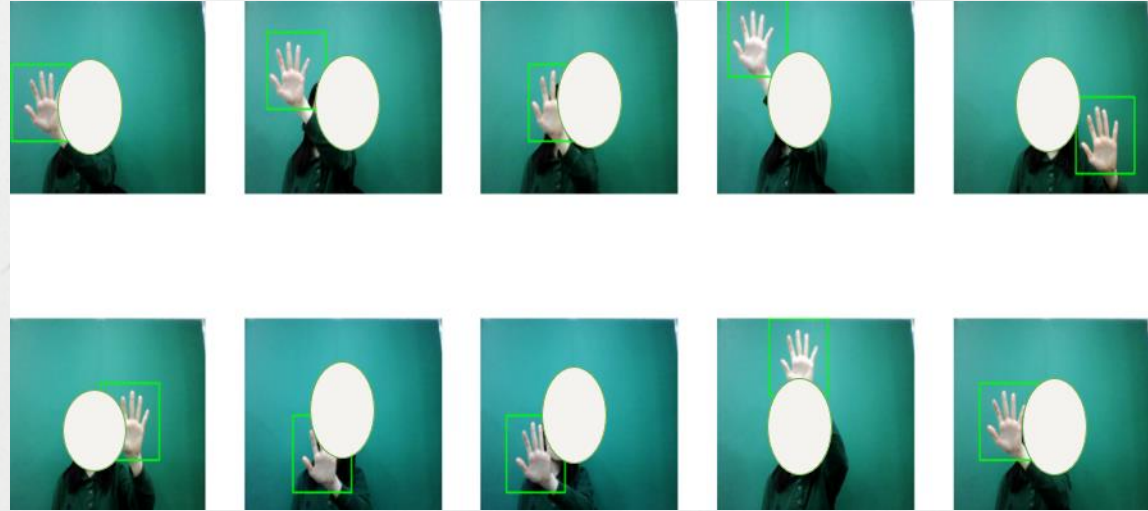
- OpenCV를 활용한 동작 감지 및 미디어 컨트롤

설명

- 음성을 통한 유튜브 영상 검색
- 손동작으로 미디어 컨트롤

적용기술

- Language: Python, CSS, HTML5, JavaScript
- OpenAPI: OpenCV, Pandas, Numpy, Google Cloud, Google Vision API
- IDE: Pycharm, Jupyter Notebook, Colab, Eclipse
- Framework: Flask, Bootstrap



Project Analysis

- Pyaudio 마이크 입력 환경 구성
- 구글 Speech To Text 환경 구축
- Youtube Data 크롤링
- Flask를 통한 웹 구현
- Object Detector 이용
- 이미지 저장 및 평가
- Hand Detector 테스트
- 제스처 컨트롤

