

데이터 분석 포트폴리오

강효창

010-4516-5162

mikkykang@gmail.com

1. 자기 소개

1-1. 이력기술서

1-2. 나를 한 단어로 표현하면?

1-3. 학업경험



회사명 텔레칩스

담당업무 웨이퍼 구매 및 외주업체 계약 관리

직위 인턴

재직 기간 2024.07 ~ 2024.08(2개월)

웨이퍼 데이터 분석 및 관리

웨이퍼 사진 데이터를 분석하여 수율이 낮은 다이의 분포를 파악하여 QA 부서에 게 전달하였습니다. 외주업체에서 웨이퍼 입고 시 LOTID별 웨이퍼 이미지 데이터를 받았고, R코드를 통해 수율이 낮은 다이는 빨간색으로 보이게 하는 알고리즘을 설계하였습니다.

해외업체와의 계약 관리

해외 외주업체와의 계약 내용을 번역 및 정리하였습니다. 자재 손실 시 물류 과정에서 양측 업체 간의 책임 범위를 얼마나 설정할 것인지, 손실에 대한 보상은 어떻게 할 것인지 등에 대하여 정리하였습니다.

MRP 기반 발주량 예측

MRP 커스텀 코드에서 자재의 당월 재고 대비 차월 수요 100% 만족하는 발주량을 예측하였습니다. 이 때 외주업체의 웨이퍼 생산 공정 데이터를 통해, 공정 완료 시점을 예측해서 추가 발주 허용량도 설정하였습니다.

1-2. 나를 한 단어로 표현하면?



중간다리

효율	<p>매 학기 어떻게 하면 짧은 시간에 좋은 성적을 낼 수 있을지 고민하였습니다.</p> <ul style="list-style-type: none">• 시험 공부 할 때 밤새서 벼락치기 하는 대신 매일 짧게 공부.• 수업 들은 직후 복습을 통한 수업 내용 장기 기억.
협력	<p>팀 프로젝트를 수행하며 원하는 결과물을 얻기 위해 항상 협력하였습니다.</p> <ul style="list-style-type: none">• 농구 동아리에 있으면서 협력에 대해 배움.• 대학원 과정 기반 팀 프로젝트를 하며 나에게 부족한 점을 메꾸고 함께 성장.
소통	<p>유연한 소통으로 팀이 전부 동의하는 결과물을 만들어냈습니다.</p> <ul style="list-style-type: none">• 내가 원하는 것과 상대방의 의견을 조율• 관계에 금이 가지 않게 서로 배려.

1-3. 학업경험

"원하는 결과물을 얻기 위한 효율적 방안 탐색"

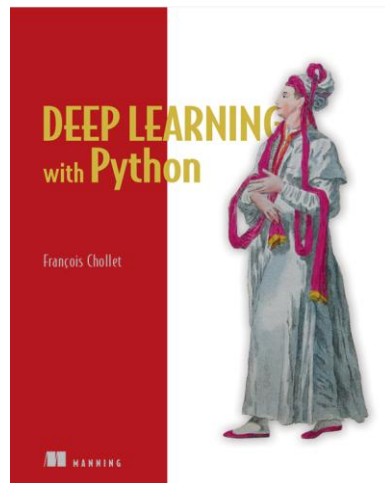
자료 찾기 능력

데이터 분석 및 시각화

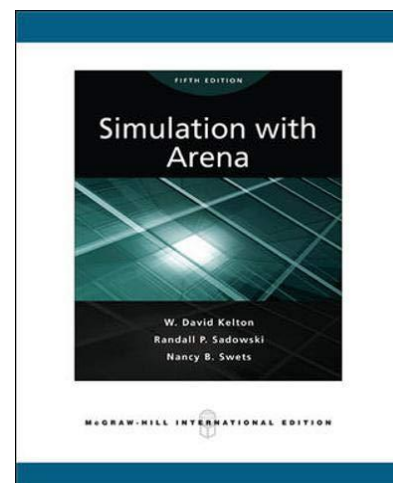
알고리즘 설계

Report

해외 교재로 수업 중 모르는 것 탐색



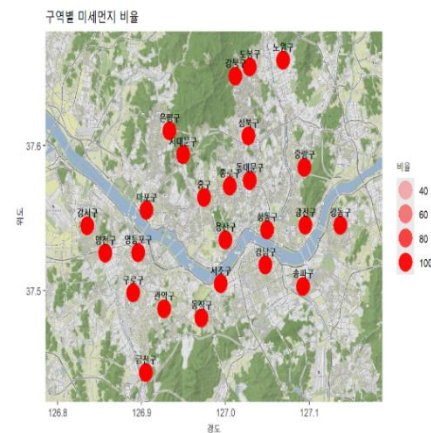
논문을 위한 딥러닝 교재



시뮬레이션 프로그램
ARENA 교재

Report

빅데이터 분석하여 패턴 발견 및 시각화



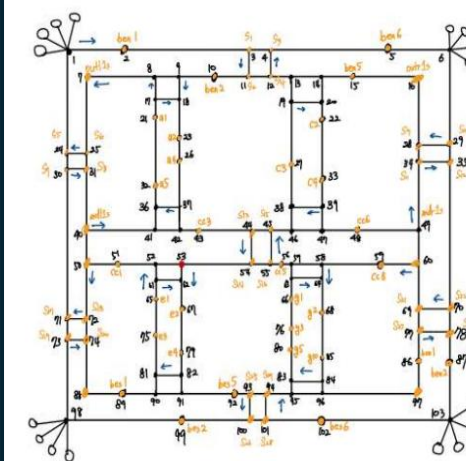
구역별 미세먼지 비율
시각화



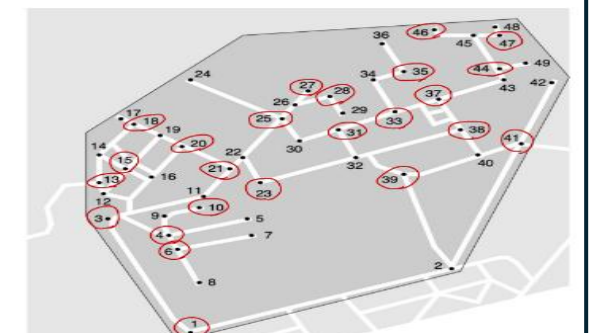
공항 이용 후기 관련
사용된 단어 분포 시각화

Report

문제 해결을 위한 효율적인 알고리즘 설계



반도체 공정에서의
OHT 효율적 배치



모든 구역 확인을 위한
CCTV 효율적 배치

2. 데이터 프로젝트

1. 신용카드 사용 금액 상환 여부 예측 모델 생성
2. 딥러닝 모델 정확도 개선 : 강아지 피부병 이미지 데이터 TFRECORD로 변환
3. 예측 모델 생성 : 공항 이용 반응에 대한 텍스트 데이터 전처리
4. 데이터 시각화 : 서울시 초미세먼지 비율 시각화

2-1. 신용카드 사용 금액 상환 여부 예측 모델 생성

	Logistic Regression	Random Forest	Decision Tree
BALANCED	46	57	65
IMBALANCED	76	76	60

	SVM Algorithm	XGBoost Algorithm
BALANCED	59	56
IMBALANCED	70	75

CLASS 균형 및 ML 알고리즘에 따라 생성된
모델의 ACCURACY

코드 : [링크](#) 논문 : [링크](#)

개발 목적 :

신용카드 사용자 데이터 전처리 후, ML 알고리즘 종류에 따른 상환여부 예측 모델의 정확도를 비교하였습니다.

상세 내용 및 후기 :

전처리 과정에서 범주형 변수를 라벨 인코딩한 후 스케일링하였습니다. 이후 종속변수의 클래스 균형 여부에 따라 머신러닝 알고리즘을 적용해 모델을 생성하고, 각 모델의 정확도를 비교하였습니다. CLASS가 BALANCED일 때가 UNBALANCED일 때보다 모델 정확도가 더 낮다는 결과를 보며, 종속변수 데이터에 대한 전처리와 주제에 적합한 데이터셋을 찾는 것이 중요함을 배웠습니다.

개발 소요 기간 :

2024.05.02~2024.06.14

사용 언어 및 패키지 :

PYTHON, PANDAS, SKLEARN, XGBOOST

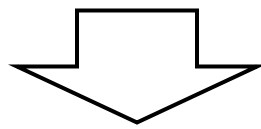
2-2. 강아지 피부병 이미지 데이터 TFRECORD로 변환

이미지 데이터



텍스트 데이터

‘우리 집 강아지가 손이 좀 붉은 것 같고
자꾸 핏대네요.’



‘무증상’
or
‘구진_플라크’

증상

코드 : [TFRECORD 생성 링크](#)(COLAB로 접속 부탁드립니다.)
[모델 생성 링크](#)

논문 : [링크](#)

개발 목적 :

AIHUB에서 제시한 이미지 데이터 기반 강아지 피부병 진단 모델에 BERT 모델을 추가하여 정확도를 개선하였음. 모델 생성 과정에서 대학원 과정인 Deep Learning 학습.

상세 내용 :

InceptionResnetV2 모델에 강아지 피부병 이미지를 FINE-TUNING한 모델과 강아지 행동에 대한 진술을 텍스트 데이터로 받아 BERT 모델에 FINE-TUNING하였음. 그리고 두 모델을 합쳐서 MULTI-INPUT 모델을 생성하였으며, 정확도는 기존 이미지 모델보다 약 20% 개선.

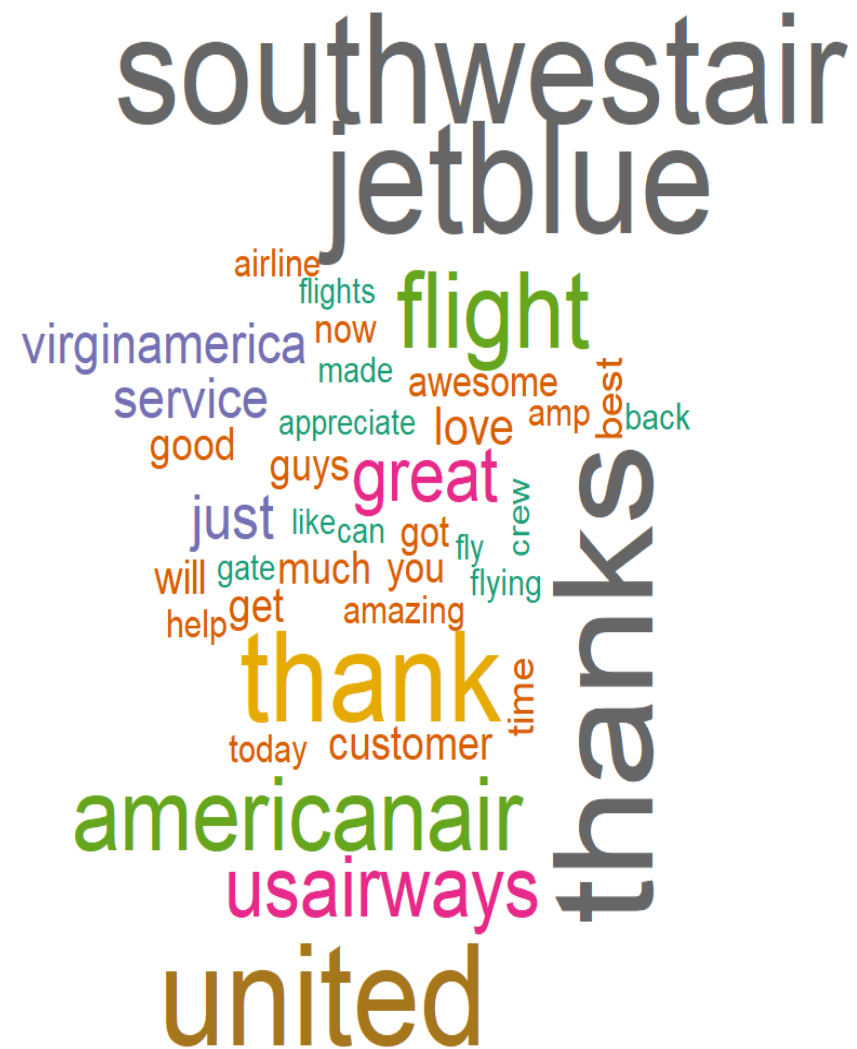
개발 소요 기간 :

2024.03.02~2024.06.14

사용 언어 및 패키지 :

PYTHON, TENSORFLOW, InceptionResnetV2, TFBertModel

2-3. 공항 이용 반응에 대한 텍스트 데이터 전처리



코드 : [링크](#)

개발 목적 :

공항 후기 관련 텍스트 데이터를 전처리한 후 단어 빈도수를 기준으로 예측 모델을 생성. 텍스트 데이터 전처리 과정 및 SVM 알고리즘을 사용하기 위한 데이터 변환 기법 학습.

상세 내용 :

먼저 텍스트 데이터를 두가지 매트릭스 형식(DTM/TF-IDF MATRIX)으로 변환한 후, 세가지 기법 (Linear, Polynomial, RBF Kernel)으로 각각 전처리후 모델을 생성. KNN 알고리즘 기반 모델을 포함하여 최종적으로 8개의 모델의 Accuracy와 F1 Score를 비교하여 적합한 모델 선정.

개발 소요 기간 :

2024/05/24~2024/06/14

사용 언어 및 패키지 :

R코드, GGPlot, WordCloud, e1071, tm

2-4. 서울시 초미세먼지 비율 시각화

코드 : [링크](#)

개발 목적 :

미세먼지 데이터를 전처리하여 초미세먼지 비율을 구한 후, 요구사항에 적합한 데이터를 시각화하여 제공

후기 :

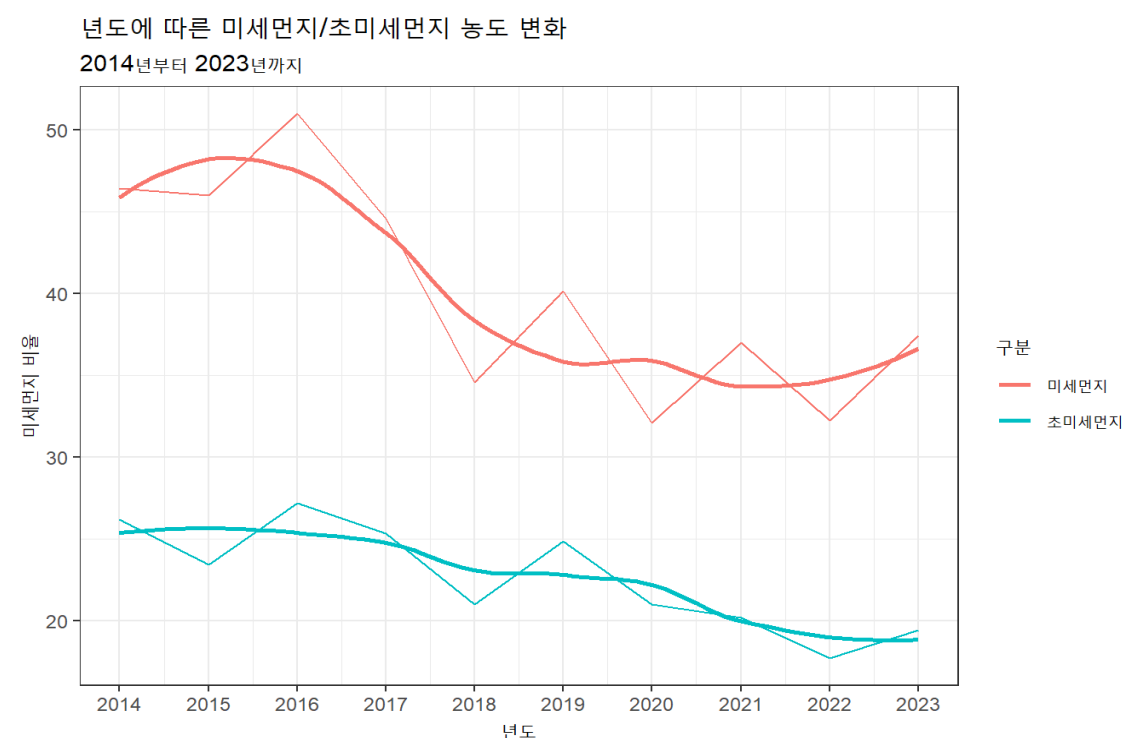
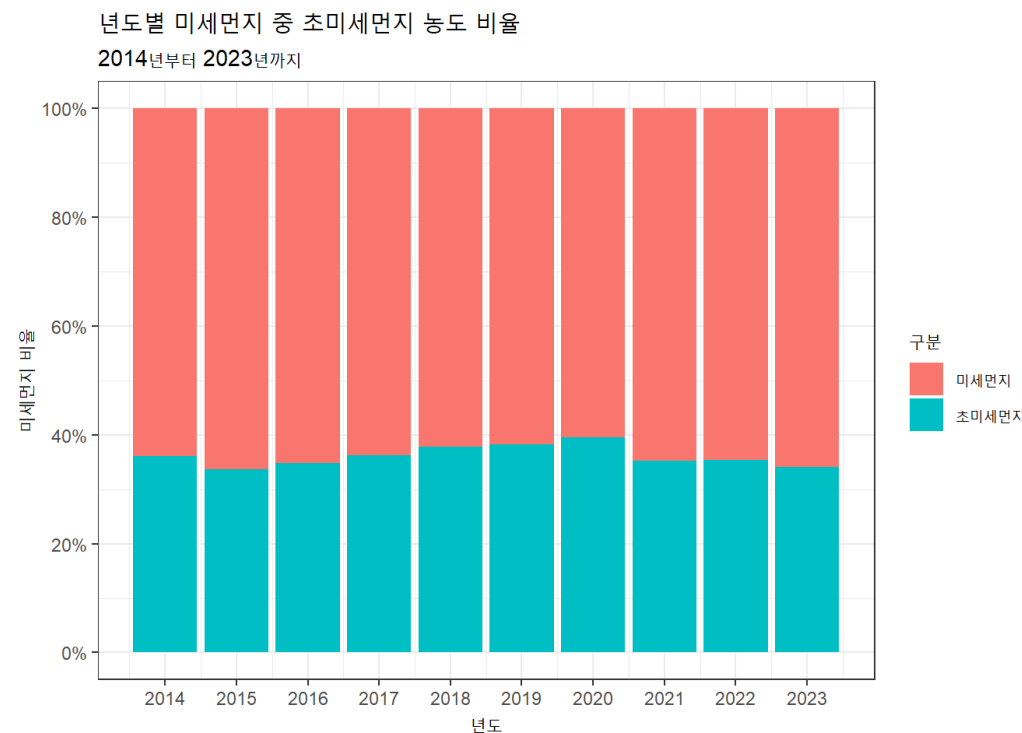
데이터를 전처리하고 정확도가 높은 모델을 생성하는 것도 중요하지만, 이를 상대방이 이해되게끔 보여주지 못한다면 의미 없는 데이터가 된다는 점을 배웠습니다.

개발 소요 기간 :

2024.03.18~2024.03.25

사용 언어 및 패키지 :

R코드, GGPLOT, GGMAP



THANK YOU

