

2019년 한이음 ICT멘토링 프로젝트 결과보고서

프로젝트명

Do You Want To be a CEO?

요 약 본

프로젝트 정보	
주제영역	<input type="checkbox"/> 건강 <input checked="" type="checkbox"/> 생산성 <input type="checkbox"/> 생활 <input type="checkbox"/> 안전 <input type="checkbox"/> 엔터테인먼트
기술분야	<input type="checkbox"/> 이동통신 <input type="checkbox"/> 방송·스마트미디어 <input checked="" type="checkbox"/> 기반SW·컴퓨팅 <input type="checkbox"/> 디지털콘텐츠 <input type="checkbox"/> 융합서비스 <input type="checkbox"/> 네트워크 <input type="checkbox"/> 전파·위성 <input checked="" type="checkbox"/> SW <input type="checkbox"/> 정보보호 <input type="checkbox"/> ICT 디바이스
달성성과	<input type="checkbox"/> 논문게재 및 포스터발표 <input type="checkbox"/> 앱등록 <input type="checkbox"/> 프로그램등록 <input type="checkbox"/> 특허 <input type="checkbox"/> 기술이전 <input type="checkbox"/> 실용화 <input type="checkbox"/> 공모전(공모전명) <input checked="" type="checkbox"/> 기타(예측 모델 완성)
프로젝트명	머신러닝 기반의 창업 위치 추천 시스템
프로젝트 소개	‘창업’이 새로운 트렌드로 자리 잡아 전문가뿐 아니라 일반인들도 창업시장에 뛰어들고 있다. 상권에 대한 이해와 분석이 필수인 창업분야에서 보다 편리하고 쉽게 상권조사를 할 수 있다면 창업을 하는 과정에서 많은 시간과 비용이 절약될 것이다. 본 프로젝트는 창업을 하는 사람들에게 시장조사의 관문을 낮춰주는 역할을 한다.
개발배경 및 필요성	창업에 관한 데이터가 무수히 많지만 새롭게 창업을 희망하는 사람들이 정보를 쉽게 얻을 수 없을뿐더러 분석하는 것이 쉽지가 않다. 그에 따라 부족한 정보를 가지고 창업을 시도하였다가 실패하는 사례들이 많아 본 프로젝트는 이러한 사람들에게 자신이 하고 싶은 업종에 대한 분석된 데이터와 해당 상권의 특징을 찾아주어 창업의 성공률을 높여 주는 것을 목적으로 한다.
프로젝트 주요기능	<ul style="list-style-type: none"> ○상권데이터 분석 <ul style="list-style-type: none"> - 서울시 여러 상권에 대한 데이터를 활용하여, 머신러닝 알고리즘을 활용 상관 관계 및 매출, 생존률, 유동인구, 유사 업종 수 등의 데이터를 분석한다. ○상권특징 분석 <ul style="list-style-type: none"> - 분석된 데이터를 활용하여 선택한 위치 주변의 상권정보를 분석하고 상권의 특징을 제공하여 창업 전 용이한 시장분석과 구체적인 계획수립에 도움을 준다. - 상권 클러스터링을 이용한 상권 세그먼트도출 및 특성 분석 ○업종 추천 및 시각화 <ul style="list-style-type: none"> - 분석된 상권 데이터를 활용하여 예비 창업자들에게 특정 위치에 창업을 할 경우 어떤 업종이 유리할지를 분석한다. - 해당 상권의 서비스 업종별로 연령대별 매출액과 상권내의 유사업종 수를 통해 상권의 뚜렷한 특징을 설명한다. - 예측 결과를 시각화하여 상권의 특징을 직관적으로 해석할 수 있다. ○새로운 상권에 대한 분석 <ul style="list-style-type: none"> - 새로운 상권이 생길 경우 해당 상권의 유동인구, 연령대별 매출액, 유사업종 수 등의 정보를 이용하여 새로운 상권의 특징을 분석한다. - 새로운 상권에 대한 세그먼트 예측
작품의 기대효과 및 활용분야	<ul style="list-style-type: none"> -예비창업자들에게 상권분석정보에 기초한 상권 및 입지분석을 제공함으로써 창업 성공률을 높이고 창업을 용이하게 한다. -어플리케이션, 웹페이지 등을 이용하여 우수한 창업역량을 가진 개인과 소상공인에게 유용한 정보를 제공한다.

(본문) 프로젝트 결과보고서

I. 프로젝트 개요

가. 프로젝트 소개

○ 기획의도: 창업을 준비하는 사람들이 창업 전 사전조사 시 알기 힘든 정보를 분석하여 쉽고 빠르게 정보를 얻을 수 있게 한다.

○ 프로젝트설명:

- 서울시 상권정보를 길 단위로 세세하게 분석하여 유동인구, 유사 업종 매출액 등 정보를 제공한다.
- 상권별 업종마다 유동인구, 상주인구, 매출액, 유사업종 수 통계량 데이터를 분석하여 다양한 업종에 대하여 자유로운 시장조사가 가능하다.

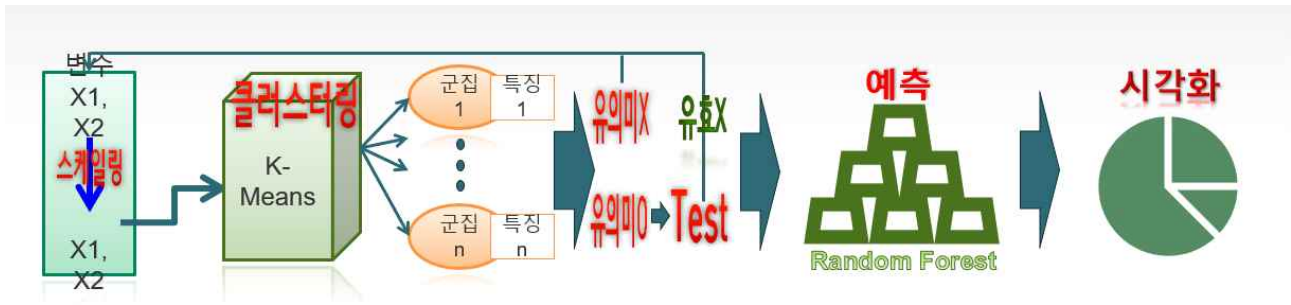
나. 개발배경 및 필요성

○ 추진배경

- 최근 한국사회에서는 취업난으로 인해 많은 인력들이 취업보다 창업에 관심을 보이고 있다.
- 처음 창업을 하는 사람들에게는 자본적 여유가 부족하므로 적은 시간과 비용으로 창업을 준비할 수 있어야 한다.
- 그러나 창업 시 필요한 정보를 얻기 위해서는 직접 발로 뛰거나 비용을 지불하고 창업지원기관에서 교육을 받아야 한다.
- 이러한 부담을 줄이기 위해 쉽게 얻을 수 없는 정보를 분석하여 사용자에게 제공함으로써 창업 부담감을 줄이고 성공률을 높일 필요성을 느꼈다.

다. 작품 구성도

0



위엔ㄴ 클러스터링으로 세그먼트 도출된 모습

그걸 가지고와서 학습해서 세그먼트 분류기를 만들었다

새로운 상권이 들어오면 세그먼트를 할당 시켜주고 시각화 독립적인걸 연결

라. 작품의 특징 및 장점

- 지도도학습을 이용하여 유동인구, 상주인구, 요일별 매출액 등을 근거로 상권이 가지고 있는 특징에 따라 프로파일을 정의하여 창업을 준비할 때 해당 상권만의 전략을 짤 수 있다.
- 지도도학습을 이용하여 새로운 상권이 생기면 통계량을 해석하여 어떤 상권에 해당하는지 분류할 수 있다.
- 연령대별 매출비율 데이터를 이용하여 상권의 분류 결과를 보면 해당 상권에 어느 연령대가 집중되어 있고, 그 연령대가 선호하는 업종이 무엇인지 분명하게 나타나므로 경쟁력 있는 창업 개념을 생각할 수 있다.
- 업종 별 통계량을 통계지표를 잘 알지 못하는 일반인들도 보기 쉽게 시각화 기법을 사용했다.

II. 프로젝트 수행결과

가. 주요기능

구분	기능	설명
S/W	상권데이터 분석	서울시 여러 상권에 대한 데이터를 활용하여, 머신러닝 알고리즘을 활용 상관관계 및 매출, 생존률, 유동인구, 임대료, 유사 업종 수 등의 데이터를 분석.
	상권 분류	분석한 통계량을 통해 상권의 특징을 비교하여 비슷한 상권끼리 모아 군집으로 분류.
	상권 예측 +시각화	상권의 통계량을 분석하여 분석해놓은 상권 군집 중 어디에 속하는지 예측.

나. 프로젝트 개발환경

구분	항목	적용내역
S/W 개발환경	OS	Windows 10
	개발환경(IDE)	Anaconda-jupyter notebook
	개발도구	notebook
	개발언어	Python

다. 장비(기자재/재료) 활용

번호	품명	작품에서의 주요기능
1	갤럭시S5	-
2	라즈베리파이	-
3		-
4		-

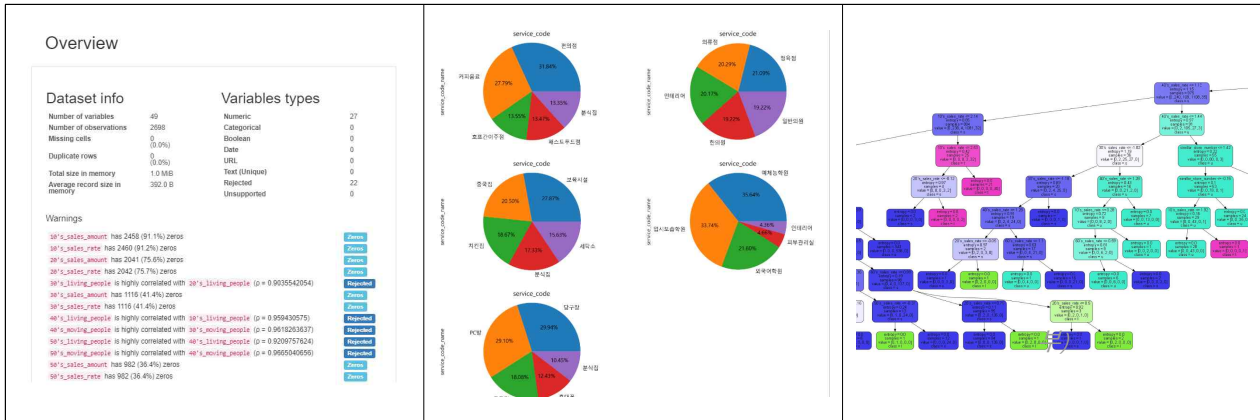
라. 프로그램 작동 동영상

#URL(유튜브에 올려서 링크 - 프로그램 작동 되는 화면을 동영상으로 작성)

0 <https://youtu.be/O4kl67qBFIA>

0

마. 결과물 상세 이미지



바. 달성성과

<input type="checkbox"/> 논문게재 및 포스터발표	게재(발표)자명	논문(포스터)명	게재(발표)처	게재(발표)일자
				2017. 00. 00.
<input type="checkbox"/> 앱(APP) 등록	등록자명	앱(APP)명	등록처	등록일자
				2017. 00. 00.
<input type="checkbox"/> 프로그램 등록	등록자명	프로그램명	등록처	등록일자
				2017. 00. 00.
<input type="checkbox"/> 특허/실용신안 출원	출원자명	특허/실용신안명	출원번호	출원일자
				2017. 00. 00.
<input type="checkbox"/> 기술이전	기술이전기업명	기술명	금액	이전일자
				2017. 00. 00.
<input type="checkbox"/> 공모전	구분(교내/대외)	공모전명	수상여부(출품/수상)	상격
<input type="checkbox"/> 실용화	#실용화한 내용에 대한 구체적 작품설명			
<input checked="" type="checkbox"/> 기타	상권의 기본 정보를 이용하여 다른 상권과 구분되는 뚜렷한 특징을 뽑아 해당 상권만의 장점, 단점을 제공하는 모델을 만들어 편리하고 직관적인 상권 분석에 도움을 준다.			

#입력한 달성성과에 대한 증빙자료는 별첨

Ⅲ. 프로젝트 수행방법

가. 업무분장

번호	성명	역할	담당업무
1	박재범	멘 토	- 지도
2		지도교수	
3	박준형	팀 장	- SW개발
4	이현수	팀 원2	- SW개발
5	박성춘	팀 원3	- SW개발
6	이유진	팀 원4	- SW개발

나. 프로젝트 수행일정

<Team 1>

구분	추진내용	수행일정								
		3월	4월	5월	6월	7월	8월	9월	10월	11월
계획	전반적인 프로젝트 이해 및 계획 흐름 설정									
분석	예비 창업자의 니즈 파악과 상권의 분석									
설계	빅데이터 전처리									
	알고리즘 및 파라미터 모형 설계									
개발	분석 알고리즘 구현									
	빅데이터 학습									
	핵심 프로그램 기능 개발									
테스트	기능을 종합하여 시범운영 및 기존서비스와 비교 후 신뢰성 확인									
종료	프로그램 시각화 및 결과보고서 작성									
오프라인 미팅	월 1회 오프라인 미팅									

<Team 2>

구분	추진내용	수행일정								
		3월	4월	5월	6월	7월	8월	9월	10월	11월
계획	전반적인 프로젝트 이해 및 계획 흐름 설정									
분석	예비 창업자의 니즈 파악과 상권의 분석									
설계	빅데이터 전처리									
	알고리즘 및 파라미터 모형 설계									
개발	분석 알고리즘 구현									
	빅데이터 학습									
	핵심 프로그램 기능 개발									
테스트	기능을 종합하여 시범운영 및 기존서비스와 비교 후 신뢰성 확인									
종료	프로그램 시각화 및 결과보고서 작성									
오프라인 미팅	월 1회 오프라인 미팅									

다. 문제점 및 해결방안

o 프로젝트 관리 측면

- 서울시 공공 데이터(우리마을 상권 분석 서비스)에 올라오는 원본 데이터에서 이상치가 있을 경우 삭제하거나 결측값을 처리하여 데이터의 신뢰성을 높여야 한다.
- 상권은 계속 변화하므로 기존에 분석해놓은 상권 클러스터를 계속 사용하는 것이 아니라 변화한 상권에 대한 정보를 분석하여 새로운 상권 클러스터를 생성하고 그 특징을 설명해야 한다.
- 빠르게 변화하는 상권에 따라 상권을 더욱 잘 설명할 수 있는 새로운 정보(X 변수)가 생길 경우 EDA와 feature selecting, 여러 알고리즘을 통한 정확도 분석으로 새로운 정보가 구체적인 상권 분석에 도움이 되는지를 확인하고 채택한다.

o 작품 개발 측면

- 원본 데이터에는 상주인구, 슈퍼 마켓 수 등 상권 및 상권 배후지에 대한 많은 정보가 있기 때문에 정확한 EDA와 feature selecting을 통해

개발 모델에서 상권을 분석할 때 유의미한 데이터인지를 파악하고
y변수와의 관계를 시각화를 통해 확인한다.

IV. 기대효과 및 활용분야

- 최근 각광받는 빅데이터 분석과 창업 환경을 접목시켜 창업의 위험부담을 감소시키고 창업 준비에 필요한 시간을 절약할 수 있게해 창업 문화를 활성화 시킨다.
- 기존 사업자들에게는 변화를 할 기반을 제공, 예비 창업자에게는 주변 상권 정보에 대한 접근성을 제공하여 창업 문화 발전에 기여한다.
- 상권 분석을 통해 소상공인의 폐업률을 낮추고 벤처기업의 생존률을 높인다.

V. 참고자료

가. 참고 및 인용자료

-
-

[별첨] 한이음 ICT멘토링 프로젝트 산출물

한이음 ICT멘토링 프로젝트 산출물

1. 분석 요구 사항 정의서 -p. 12
2. 분석 데이터 정의서 -p. 13
3. 분석 모형 정의서 -p. 15
4. 분석 정의서 -p. 16
5. 전체 분석 시스템 구성도 -p. 20

#소스코드 등 프로젝트 수행 중 발생한 산출물 및 달성성과 증빙 첨부

<프로젝트 산출물>

멘토 박재범

멘티 홍익대학교 컴퓨터공학과 박준형, 박성춘, 이현수, 이유진

요약

창업 시 필수적으로 필요한 시장조사를 편리하게 하기 위해 머신러닝 기법을 이용하여 상권을 분류하고 예측한다. 전체 모델은 Team 1(현수, 유진), Team 2(준형, 성춘) 두 팀으로 나누어 진행한다.

Team 2는 상권 데이터를 활용하여 K-means 클러스터링으로 클러스터를 생성하고 상권을 특징별로 grouping한다.

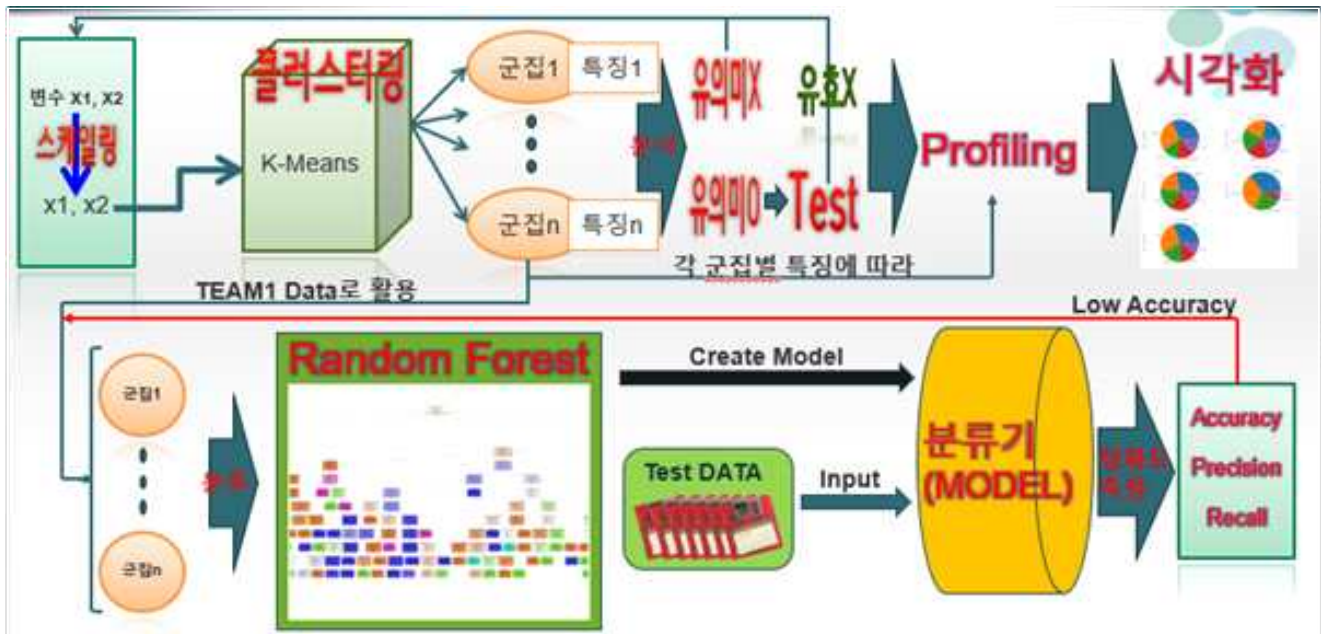
Team 1은 Team 2에서 결과로 나온 클러스터 Data를 Input으로 Random Forest를 사용하여 분류기(Model)를 만든다. 새로운 상권이 Input으로 들어왔을 때 어느 클러스터에 해당하는지 분류하는 기능을 한다. 이 모델은 클러스터 별 특징과 분류된 클러스터에 해당하는 업종을 통해 편리한 상권 분석을 할 수 있도록 도와준다.

1. 분석 요구 사항 정의서

- 상권 분석에 대한 요구사항
 - 상권을 새로운 특성으로 분류한 후 Grouping이 필요하여 Clustering
 - 분류한 cluster에 대해서는 명확한 근거 필요
 - 새로운 상권이 어느 cluster에 해당하는지 분류 및 예측
 - 새로 입점 시 해당 상권의 정보를 분석, 예측

2. 전체 분석 시스템 구성도

전체 시스템을 구성도로 표현하면 아래와 같다.



3. 분석 데이터 정의서

3-1. 데이터 정의서

- 데이터의 출처
 - 서울 열린 데이터 광장 - 서울시 우리마을가게 상권분석서비스
- 데이터 항목과 건수

데이터 원본 이름	기간 및 건수	항목
서울시 우리마을가게 상권분석서비스 (상권-추정유동인구)	2018년 9, 10월	총 유동인구 수, 연령대 별 유동인구 수
서울시 우리마을가게 상권분석서비스 (상권-상주인구)	2016년 4월 2018년 10월	총 상주인구 수, 연령대 별 상주인구 수

서울시 우리마을가게 상권분석서비스 (상권-추정매출)	2018년 9, 10월	평균 영업 개월 수, 점포 수
서울시 우리마을가게 상권분석서비스 (상권-점포)	2018년 10월	점포 수, 유사 업종 점포 수, 개업율, 개업 점포 수, 폐업율, 폐업 점포 수
서울시 우리마을가게 상권분석서비스 (상권-지수지표)	2018년 8, 9, 10월	상권 안정성, 활성화도, 성장성, 지수/지표 구간

3-2. 최종 변수 정의서

2-1의 데이터 중 최종적으로 모델에 활용된 항목을 팀별로 입력 변수(X) 와 결과 변수(Y)로 나누어 정의한다. Team 2의 클러스터링 결과로 나온 클러스터(Y)를 Team 1에서 X변수로 활용하여 새로운 상권 데이터가 들어왔을 때 Team 2의 Y 값 중 어떤 클러스터에 해당하는 지를 결과값으로 예측한다.

■ Team 2

	변수명	변수의 의미
X변수 각 연령층의 매출비율과 유사업종수로 클러스터링	X1: 10's_sales_rate	10대 연령층의 매출 비율
	X2: 20's_sales_rate	20대 연령층의 매출 비율
	X3: 30's_sales_rate	30대 연령층의 매출 비율
	X4: 40's_sales_rate	40대 연령층의 매출 비율
	X5: 50's_sales_rate	50대 연령층의 매출 비율
	X6: 60's_sales_rate	60대 연령층의 매출 비율
	X7: silmilar_store_number	유사업종 수
Y변수	Cluster_id	각 군집에 해당하는 번호

■ Team 1

	변수명	변수의 의미
X변수 각 연령층의 매출비율과 유사업종수로 클러스터링	X1: 10's_sales_rate	10대 연령층의 매출 비율
	X2: 20's_sales_rate	20대 연령층의 매출 비율
	X3: 30's_sales_rate	30대 연령층의 매출 비율
	X4: 40's_sales_rate	40대 연령층의 매출 비율
	X5: 50's_sales_rate	50대 연령층의 매출 비율
	X6: 60's_sales_rate	60대 연령층의 매출 비율
	X7: silmilar_store_number	유사업종 수
Y변수	Cluster_id	각 군집에 해당하는 번호

3-3. Scaling

종류	설명
StandardScaler	기본 스케일, 평균과 표준편차 사용
MinMaxScaler	최대/최소값이 각각 1, 0이 되도록 스케일링
MaxAbsScaler	최대절대값과 0이 각각 1,0이 되도록 스케일링
RobustScaler	중앙값과 IQR 사용. 아웃라이어의 영향을 최소화

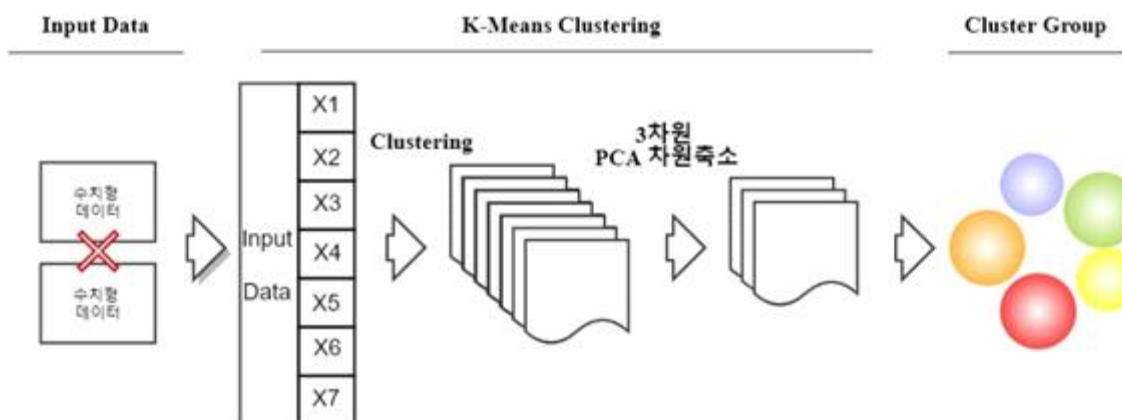
- ✓ 기본 데이터의 값 범위가 다양하기 때문에 일부 기계 학습 분류 알고리즘에서 목표 함수가 정규화없이 제대로 작동하지 않음
- ✓ 알고리즘에서 두 노드 사이의 거리가 유클리드 거리로 계산하는데, 광 범위 한 경우 정확도 및 오차율이 커지는 경향이 있어 범위를 표준화 해주는 작업이 필요
- 평균과 표준편차를 사용한 StandardScaler 사용

4. 분석 모형 정의서

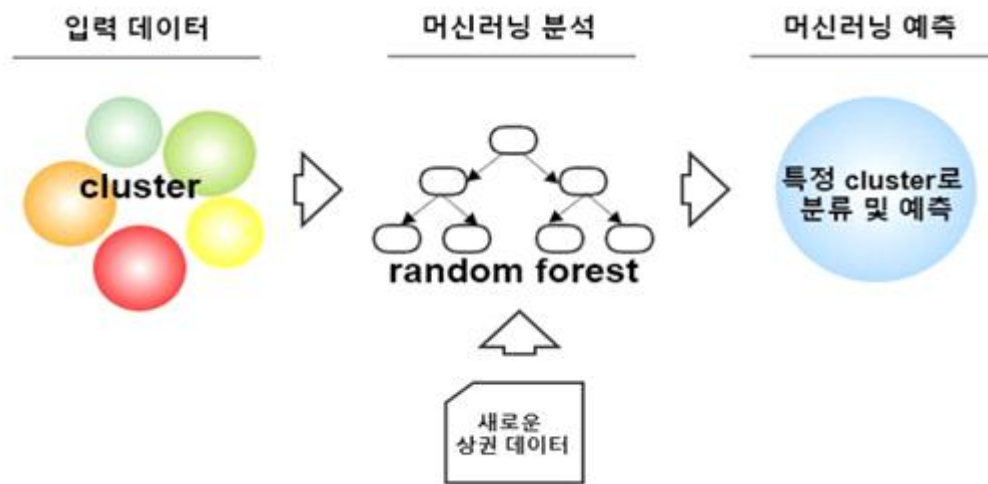
- 데이터 수집 -> 탐색적 분석 -> 변수 정의 -> 알고리즘 적용 ->
-> 학습 -> 비지도 학습: 해석, 지도 학습: 모형 정확도 해석 -> 시각화

4-1. 팀별 모형

◆ 클러스터링 모형(team 2)



◆ 예측 모형(team 1)



5. 분석 정의서

◆ 비지도 학습: Clustering (Team 2)

Scikit learn의 K-means clustering을 이용하여 클러스터링을 한 결과를 분석한 클러스터 별 의미는 아래와 같다.

5.1 군집 분석(clusters Profiling)

강조 - 가장 큰 값

K=5	Group1	Group2	Group3	Group4	Group5
Group Size	2690	11696	1066	13724	2371
10's_sales_rate	3.189169e-01	5.529788e-01	1.923211e+01	2.157160e+00	1.370154e-01
20's_sales_rate	2.223999e+00	7.101159e+00	3.681662e+01	3.165609e+01	6.476657e+00
30's_sales_rate	1.032489e+01	1.316852e+01	1.185722e+01	2.482802e+01	6.171897e+01
40's_sales_rate	7.286105e+01	2.112593e+01	1.573173e+01	2.044646e+01	2.229454e+01
50's_sales_rate	1.114841e+01	3.163107e+01	1.174457e+01	1.422327e+01	6.058024e+00
60's_sales_rate	3.122255e+00	2.642082e+01	4.617092e+00	6.689407e+00	3.315006e+00
act_jipyo_value	5.101726e+01	5.102993e+01	5.238094e+01	5.171305e+01	5.031214e+01
growth_jipyo_value	5.415332e+01	5.244782e+01	5.660254e+01	5.272788e+01	5.283213e+01
safety_jipyo_value	5.511827e+01	5.508823e+01	5.429945e+01	5.412280e+01	5.506596e+01
total_moving_people	4.992590e+04	5.178091e+04	5.981758e+04	5.709191e+04	5.158410e+04
total_living_people	3.093916e+01	3.010790e+01	2.836930e+01	2.821042e+01	2.930059e+01
c_month_sales_amount	3.219436e+07	6.769874e+07	2.870573e+07	6.178005e+07	2.305094e+07
Rank of Service (Top5)	1. 예제능 (617)	1. 정육점 (732)	1. PC방 (149)	1. 편의점 (1206)	1. 보육시설 (848)
	2. 임시보습 (557)	2. 의류점 (696)	2. 당구장 (126)	2. 커피음료 (1117)	2. 예제능학원 (109)
	3. 외국어학원 (363)	3. 인테리어 (692)	3. 문구점 (78)	3. 분식집 (825)	3. 피부관리실 (99)
	4. 피부관리 (87)	4. 일반의원 (676)	4. 분식집 (53)	4. 치킨집 (746)	4. 세탁소 (76)
	5. 화장품 (76)	5. 약국 (668)	5. 휴대폰 (49)	5. 중국집 (703)	5. 화장품 (76)

5.2 각 Group 특징

- ☞ Group(1): 안정성이 제일 높고 상주인구 값이 제일 높다.
직종의 대부분이 학원이라는 특징을 갖고 있어 주 매출 연령대가 부모 연령대인 40, 50대 매출비율이 높다. (학원가 상권의 특징)
- ☞ Group(2): 평균 영업 개월 수가 가장 높고 40~60대의 매출비율이 높게 나타난다.
대체적으로 의류, 약국, 병원 같은 생활 밀접 업종들이 다량 분포.
(시장 상권의 특징)
- ☞ Group(3): 10대, 20대의 매출비율이 특히 높으며 유동 인구가 많은 지역들을 나타내고 있다.
그러나 유동인구에 비해 총 매출액이 그다지 높지 않다.
활성도 지표가 제일 높음. (젊음의 거리)
- ☞ Group(4): 20~40대까지의 매출 비율이 비슷하며 상주인구에 비해 유동인구가 높다.
Group3과 비슷한 성향을 갖고 있지만 월 평균매출량이 가장 높다는 점이 다름.
대체적으로 편의점, 카페, 호프집과 같은 유흥 상권이 형성 되어있다.
(일반 변화가 상권)
- ☞ Group(5): 보육시설이 주로 위치해 있기 때문에 월 평균 매출이 가장 낮다.
주 매출 연령대는 부모연령대인 30대이다.
(젊은 부부 상권)

5.3 Labeling

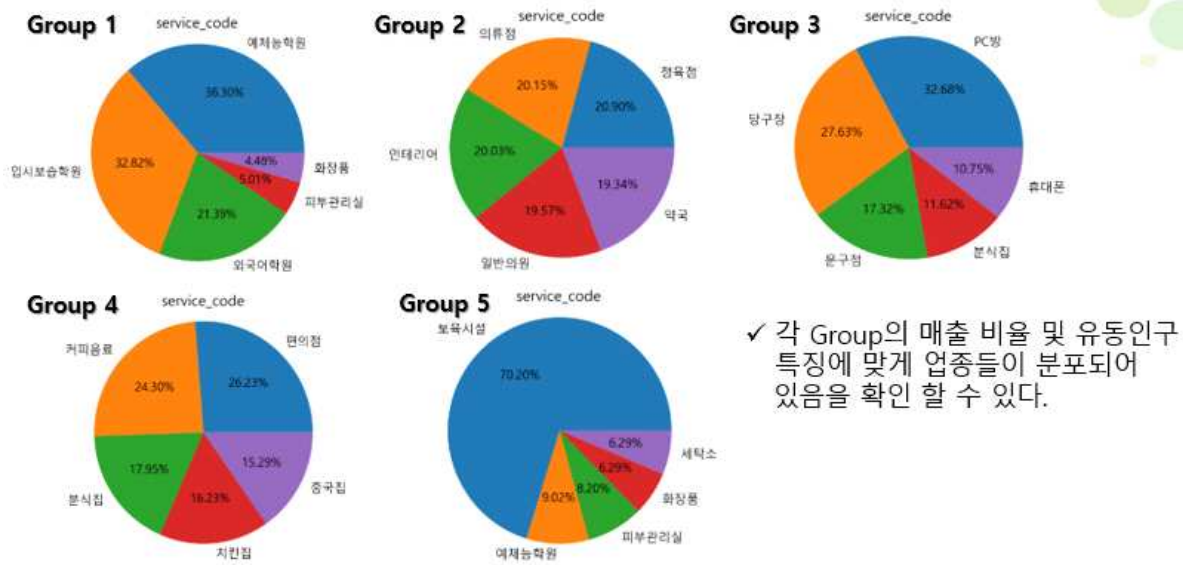
- Group(1): 학원가
- Group(2): 시장 상권
- Group(3): 젊음의 거리
- Group(4): 일반 변화가 상권
- Group(5): 젊은 부부 상권

5.4 K-means Clustering 기법: K(군집의 수)값의 변화에 따른 군집 유형 분석

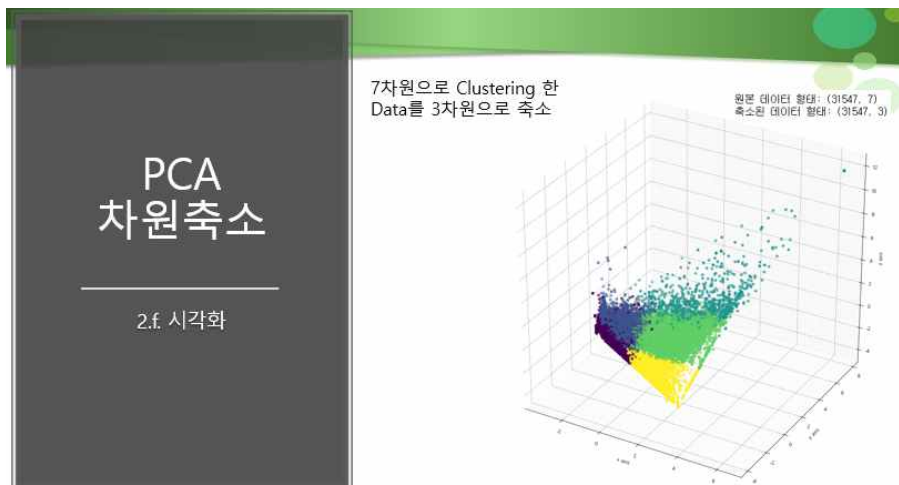
K의 크기 (군집의수)	K에 따른 군집 유형 분석
K=3	군집의크기가 약 12000, 15000, 3000 개로 정확한 군집 별 특징을 파악하기에 어려움이 존재
K=5	군집의 크기가 약 2500, 2000, 13000, 12000, 1000 개로 나뉘었는데 군집 별 특징을 파악 가능
K=7 (1)	몇몇 군집에 속한 데이터의 개수가 너무 적어 유의미하다 보기 힘들
K=7 (2) K=3으로 clustering 후 Group의 크기가 큰 Group들을 K=3 Clustering	약 12000, 15000 개로 나뉜 집단을 같은 조건(K=3)으로 clustering 한 결과 군집 별 데이터의 개수가 더 편향적으로 나타나는 현상

결론: K=5 일 때 군집 별 크기의 편차가 작고 특징이 뚜렷해서 K=5 사용

5.5 시각화



5.6 PCA 차원축소



◆ 지도 학습: 예측 정확도 (Team 1)

Random Forest를 이용하여 분류한 결과인 평가지표별 지표 값과 정오 분류표 (Confusion Matrix)는 아래와 같다.

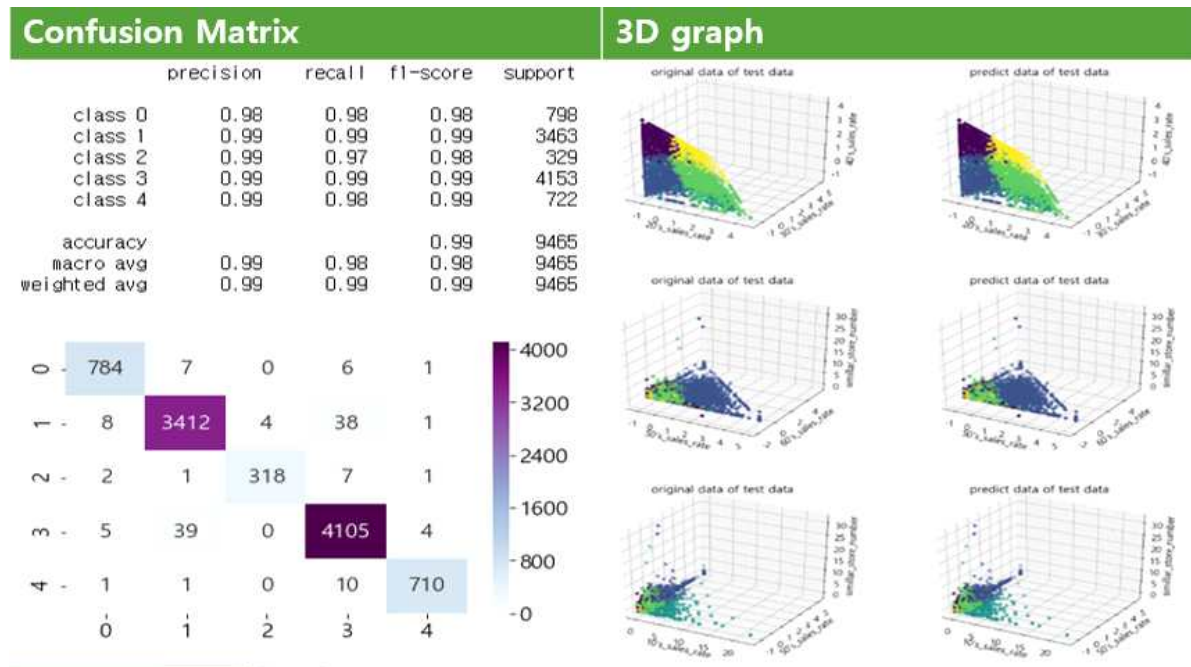
5.7 평가지표별 지표값

지표 유형	지표	지표설명 및 값	
정/오탐 유형	TP(True Positive)	A를 A로 정 탐지	
	TN(True Negative)	B를 B로 정 탐지	
	FP(False Positive)	B를 A로 오 탐지	
	FN(False Negative)	A를 B로 오 탐지	
예측 정확도	Accuracy = (TP+TN)/(TP+TN+FP+FN)	0.986	
재현율	Recall = (TP)/(TP+FN)	Class 0	0.98
		Class 1	0.99
		Class 2	0.97
		Class 3	0.99
		Class 4	0.98
정밀도	Precision = (TP)/(TP+FP)	Class 0	0.98
		Class 1	0.99
		Class 2	0.99
		Class 3	0.99
		Class 4	0.99

5.8 Confusion Matrix

Confusion Matrix		예측값					
		Class 0	Class 1	Class 2	Class 3	Class 4	합계
실제값	Class 0	784	7	0	6	1	798
	Class 1	8	3412	4	38	1	3463
	Class 2	2	1	318	7	1	329
	Class 3	5	39	0	4105	4	4153
	Class 4	1	1	0	10	710	722
	합계	800	3460	322	4166	717	9465

5.9 시각화



6. 전체 코드

