05.16 회의록

질의사항

1. sql 사용 질의

db에 저장하면 연동하는 것을 할 줄아냐, 우리 배운 것은 ipynb로 불러 올 수 있나

정형화된 데이터가 있다면 사용

유튜브는 제외

이번에는 머신러닝이 들어가는데, 크롤링 데이터 중 숫자 데이터를 분석하는데,

감성분석은 0또는 1로 나눠지는데, 그래서 그냥 분석으로 변경

sql은 정형 데이터 분석

상관이 없다면 안해도 되는데, 정형데이터가 있다면 sql 이용

-> 쓰면 좋고 안써도 무방(쓰는 것을 선호)

텍스트 마이닝을 하게되면 빈도 수 등을 기초통계 가능?

LDA 기법

1. 자연어 처리 부분

-> 강사님 다른 방안 제시

텍스트마이닝 기법으로 데이터를 수치화하여 하기

수치화해야 머신러닝이 가능

* 정형데이터 분석
* 비정형데이터 분석

텍스트 마이닝: 어떤 핵심단어가 제일 많이 언급되었는지 등을 뽑을 수 있다

댓글 분석만 하게된다면, 결과가 초라할 수 있음

텍스트 마이닝을 사용하면 확률값으로 앞에 나온 단어로 뒤에 단어를 예측할 수 있다.

이는, SQL로 끌어올 수 없다

현실화…..?

평가 지표

* 함수, 모델링(머신러닝), 기초통계

방향성 변경

* 주제 관련해서 어떤 숫자값을 이용할 수 있는지
* 회귀나 머신러닝을 사용할 수 있는 주제로 변경
* 현재 주제에서 정형데이터로 나와야 한다면, 같은 주제로 선출하는 내용들이 빈번
* 채널 별로 조회수나 댓글의 핵심 단어 빈도수

-> 주제에 대한 빈도수체크를 하며 기간 별 사람들은 어떤 단어를 많이 사용했는지

-> 뉴스에서 어떤 뉴스를

숫자 데이터를 통계분석과 머신러닝을 사용할 수 있는지

머신러닝을 기준으로 무엇을 예측할 수 있는지

-> 유튜브에서 숫자로 얻을 수 있는거? 좋아요수 , 구독자수, 댓글수, 댓글 길이 수

숙박 시설은 별점으로 해볼 수 있지 않을까?

숫자 데이터를 많이 확보할 수 있는 주제를 조금 찾아보면 어떨까?

강사님의 생각? 여행패키지를 고를때 소비자의 기준을 댓글 분석으로 파악하여

선호도를 수치화하여 추천 시스템

주제 댓글 리뷰

댓글을 통계적 기법으로 주제를 분석?을 넘어

다음 것을 예측하는 서비스를 제공

의견

* 게스트 하우스를 이용하는 사람들에게 추천 서비스
* 추천시스템 (사람의 취향을 파악할 수 이는 요소를 사용자 리뷰에 빈도수가 많은 것을 엮어서 추천)
* 댓글을 후반 부속물로 두고, 월세 동향에 대해 파악하고 정하는데, 부동산 쪽으로 월세 동향이 어떻게 되는지 크롤링 후 댓글 분석 후

서울로 취업하는 첫 집 추천 시스템

댓글을 메인으로 할 수 없다는 점 / 월세 동향이 메인

입지 선정인데-> 주변 상권등을 파악하여

주 문제점은 댓글

* 신혼부부로 타겟을 한다면? 호갱노노나 집품에 월룸에 비해서는 댓글이 많음

집의 특성을 리뷰와 우리의 기준으로 평점을 만든다.

그 점수의 경우 객관적인 지표가 가능한것인지에 대한 문제

그 가중치를 어떻게 줄 것인지에 대해

굳이 가중치를 주지 않더라도, 집 주변에 몇개가 있는지에 대한

순위를

* 스마트 기기의 경우 댓글이 깔끔하다는 장점이 있다.
* 외국 데이터로 보면은 숫자형을 찾을 수 있다.
* 예측을 기본으로 해야하는데 텍스트 분석이 가장 문제다…………………………….

텍스트 분석 말고 사용자 리뷰에는 평점, 그외 다른거?

예측 아니면 분류 -> 고객 세분화?

어떤 상품 또는 장소의 리뷰를 다 분석하여 빈출 단어를 파악하여 사용자를 적응할 수 있다면, 어떤 니즈를 가지고 있는지, 분류하여 고객을 세분화하여

이러한 고객은 이러한 마케팅 컨설팅을 제시

분류 기법이 많음 ->

현대백화점 체류시간 / 소비 금액 / 등등

분류하여

-

뉴스의 경우

* 텍스트 분석

상품마다 고객 비율이 다름 충성도가 높은 사람을 찾을 수 있다면,

쿠팡에서 무작위로 10개를 선정하고

재구매율이 높고 충성도가 높은 고객이 있다.

충성고객(재구매 할 의향이 있는 고객)과 이탈 위협 고객으로 나누고 데콘에 있는 데이터로 고객 세분화하고 맞춤 솔류션을 도출하고, 실제 제품에 어떻게 적용시킬 수 있는지

재구매율을 예측하는 모델

별점과의 상관관계

상품에 대한 재구매자 예측과

피쳐

* 가격
* 별점
* 재구매율
* 만족도

소모성과 기능성이 있는

화장품이라면 화해

이 다음에는 어떤 제품이 좋을 것인지 제안, 개선

분류 회귀랑 예측 회귀를 사용할 수 있다………..

재구매 몇개

피쳐로 평점 연령대 성별을 적용하여 범주화되어 있고,

피부 타입 , 발림성, 자극도

제품 추천 또는 개선 방향

샘플을 재구매하는 고객으로

Y = 재구매

X =

이런 피쳐들

걱정 되는 것이 논리적으로 연결성이 약할 수도 있다.

-> 댓글 재구매율

샘플을

궁극적으로

화장품 제품마다 특징이 있다.

해당 제품의 1점 2점대의 리뷰를 분석하여

피부 타입 건성 복합 지성 으로 카테고리를 설정하고

발림성 등과 같은 피쳐는

강한 특성을 가진 피쳐를 가진 주제이어야 함.