Big Data X Korea Univ.

<<빅데이터 프로젝트>>

**[11조]**

**팀원 : 고민정, 김예린,**

**서민지, 신동윤, 정윤지**

목 차

1. 프로젝트 주제 3

1.1 주요 키워드 3

2. 각 프로젝트별 상세사항 3

2.1 도서관 시설물 정보 3

2.1.1 문제 인식 3

2.1.2 목표 3

2.1.3 사용 데이터 3

2.2 그 외의 도서관 관련 제공 서비스 3

2.2.1 도서 추천 : 도서관, 사이트 연결 / 대출가능한 도서 위주 추천 3

2.2.2 책 검색 : 도서관 찾아주기 3

2.2.3 커뮤니티 : 서로 소장한 도서목록 공유 (공유 경제 개념) 3

2.2.4 전공책 공유 서비스에 초점 3

2.3 경마에서 발생하는 데이터를 통한 승부 예측 및 배팅 전략 수립 4

2.3.1 문제 인식 4

2.3.2 목표 4

2.3.3 사용 데이터 4

2.4 창업자를 위한 상권 분석 시스템 4

2.4.1 문제 인식 4

2.4.2 목표 4

2.4.3 사용 데이터 4

2.5 기계학습을 통한 유가 예측모델 4

2.5.1 문제 인식 4

2.5.2 목표 4

2.5.3 사용 데이터 4

2.5.4 요약 5

2.6 스마트 팜 작물 추천과 환경제어 서비스 5

2.6.1 문제 인식 5

2.6.2 목표 5

2.6.3 사용 데이터 5

2.6.4 요약 5

3. 추가사항 5

3.1 0719 회의록 5

# 프로젝트 주제

주요 키워드

: 도서관, 경마, 창업, 스마트 기름, 스마트팜

# 각 프로젝트별 상세사항

도서관 시설물 정보

문제 인식

: 도서관을 이용할 때, 사람마다 도서관 시설물에 대한 선호도가 다르다.

목표

: SNS상의 도서관 시설에 대한 정보를 텍스트 마이닝하여, 키워드를 뽑아 분석하는데 이용하고자 한다.

사용 데이터

* 각 도서관 위치 정보
* 각 도서관 시설물 정보 ex)내부 시설, 수용 인원
* 각 도서관 이용자 수 정보
* 각 도서관 대출 기록
* SNS 도서관 후기 정보

그 외의 도서관 관련 제공 서비스

도서 추천 : 도서관, 사이트 연결 / 대출가능한 도서 위주 추천

책 검색 : 도서관 찾아주기

커뮤니티 : 서로 소장한 도서목록 공유 (공유 경제 개념)

전공책 공유 서비스에 초점

경마에서 발생하는 데이터를 통한 승부 예측 및 배팅 전략 수립

문제 인식

: 경마 승부를 효율적으로 예측하기에 어려움이 있다.

목표

: 경마에서 발생하는 데이터를 통한 승부예측 및 배팅 전략 수립하고자 한다. 또한, 단순 승부 예측 외에도 승부조작 방지 등에 활용 가능할 수 있을거라 생각한다.

사용 데이터

* 경기마 정보
* 우승자 정보
* 날씨 정보
* 경기 장소 정보
* 우승 전적

창업자를 위한 상권 분석 시스템

문제 인식

: 창업 시에 시장 분석, 소비자 분석을 위한 신뢰성 있는 정보가 필요하다.

목표

: 위치, 업종을 선택하면 주변 경쟁업체, 유동인구, 지역 인구 특성 등을 분석하여 해당 창업자에 알맞은 상권을 추천한다.

사용 데이터

* 각 지역별 상권 정보
* 시간별 유동인구 정보
* 현재 트렌드 정보(SNS crawling)

기계학습을 통한 유가 예측모델

문제 인식

: 유가 예측은 정부와 관련 업계에게 중요한 문제이다.

목표

: 정확성과 정밀도가 높은 유가 예측 모델을 구축한다.

사용 데이터

* 과거 유가 정보
* 세계 기후 협약 동태
* 유가에 대한 국제적 여론 정보
* 대체제, 보안제 현황 정보

요약

: 유가를 예측할 때, 보통 예측을 위해 시계열 기법을 사용하지만 유가의 흐름은 정량적으로 설명되지 않는 여러 예측되지 않고 불규칙한 외부 요인에 의해 좌우된다. 따라서 예측모델에서 이런 외부 요인들을 제거하지 않고 쌓인 데이터들을 통해 학습시키고 훈련시켜 다른 모델보다 정확성과 정밀도가 높은 모델을 구축한다.

스마트 팜 작물 추천과 환경제어 서비스

문제 인식

: 국가에서 주도하고 있는 농업 증진 과제의 중심에 있는 ‘스마트 팜’의

현재는 작물의 종류가 26가지로 한정되어 있다.

목표

: 국가의 농업활성화를 위하여 농장을 새로 증설하고자 하는 사람들에게 현재의 트랜드에 맞는 작물들을 위치, 날씨 등을 고려하고 추천해주고, 이에 따른 개인농장주들에게 customized화된 농장의 환경을 제공하고자 한다.

사용 데이터

* 스마트팜 환경 정보
* 작물 생육 정보
* 영농 경영 정보

요약

: 농작물에 영향을 끼치는 관련 기사를 crawling 해서 키워드를 뽑아내고, 텍스트 마이닝 변수로 집어넣는다. 경영정보, 작물정보, 다른 데이터, 위치정보(작물이 자라기 좋은 환경), 수도, 전력과 같이 비용을 최소화하는 것을 목적으로 계절별, 분기별 농작물을 추천해준다.

기존에 있는 환경데이터 가지고 학습을 시켜서 자동으로 조절해 줄 수 있는 환경을 시스템으로 실시간으로 보여주는 것을 목표로 하고 있다.

# 추가사항

0719 회의록

<회의 내용>

* 연체 데이터 분석해 대출 정책 수립
* 출판 예정인 도서의 수익성 분석
* 도서관 데이터를 ‘목적’이 아닌 ‘수단’으로 쓰자
* 도서 대출 데이터를 분석해 지역별 특성을 도출하여 타 분야의 예측시스템에 활용해보자. (도서관 데이터를 변수로 활용)
* 도출 가능 내용 : 지역별 인구 유형과 구성, 관심분야
* 도서관 데이터를 분석해 주변 상권 부동산 시세 예측
* 특허 도서관 : 특정 주제에 집중한 도서관
* 대출 자료를 통해 특허도서관의 주제를 추천
* 도서관 ‘잉여 자산’에 대해 생각해보자
* 문화 프로그램 : 관심 저조
* 이용률이 저조한 남는 시설
* 등등 위의 잉여자산을 이용하여 수익성을 낼 수 있는 방법이 없을까?

<회의 결론>

* 고민 중인 내용( ‘도서관’ 주제의 문제점과 세부사항)을 기획서에 추가하여 제출하고 피드백을 기다리기로 결정
* 피드백 받기 전까지 각자 주제에 대해 자유롭게 생각해올 것