



Term Project 소개

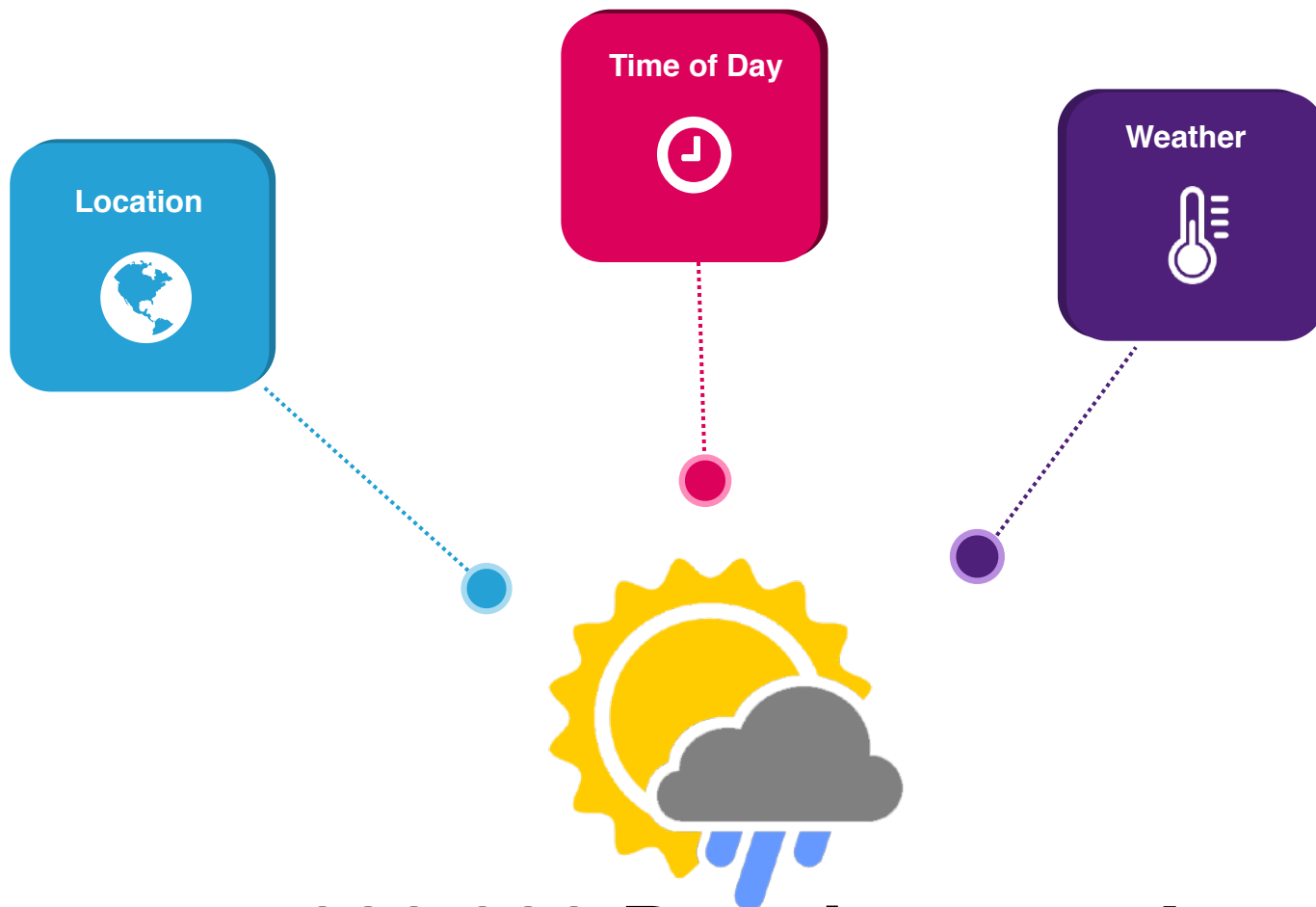
2017-1 데이터마이닝 프로젝트

연세대학교 컴퓨터과학과
소프트컴퓨팅 연구실

프로젝트 개요

- 주제
 - 기상청 날씨 데이터 대상
 - 새로운 가치 창출: 일상생활의 유익성과 비즈니스 수익성 증대
- 목적
 - 수업 이해도: 빅데이터에 대해 데이터 마이닝 기술을 적용
 - 실무적 경험: 실생활에서 발생한 데이터로부터 유의미한 분석 도출 경험

기상청 날씨 데이터



5,000,000 Data instance!
(for 6 years)

- Attribute

Time

- 4

기상청 날씨 데이터 속성

| | Feature name | | Feature name |
|------------|---------------------------|------------|-------------------------|
| Metadata | Region | Obervation | Snow cover |
| | Date(Time stamp) | | Snow cover(3h) |
| Obervation | Temperature | | Sky coverage |
| | Rainfall | | Sky coverage(mid) |
| | Wind velocity | | Cloud shape |
| | Wind direction | | Cloud height(mx-mn) |
| | Humidity | | Visibility |
| | Atmospheric pressure | | Ground condition |
| | Dew point | | Phenomenon index |
| | Spot atmospheric pressure | | Ground temperature |
| | Sea-level pressure | | Under(5cm) temperature |
| | Sunshine amount | | Under(10cm) temperature |
| | Insolation | | Under(20cm) temperature |
| | | | Under(30cm) temperature |

- 27개 속성
- 관측치가 없는 경우 Null
 - 적설량
 - 일조
 - 일사
 - 운량
- Sparse

기상청 날씨 데이터 분석 예시

- 2016경연대회 최우수상(치킨-알파고팀):
"주말기온이 25 °C를 넘으면 20~40대 여성들의 치킨주문이 많아진다"
 - 기상자료와 이동통신사 배달음식 주문데이터 모델링
 - 날씨에 따라 변화하는 치킨과 중국음식의 주문량을 예측
 - 자영업자들이 현실적으로 활용할 수 있는 방안을 제시
- 2015경연대회 최우수상:
"날씨로 인한 문화관련 행사 취소 담보 보험-웨더CARE"
 - 날짜, 위치에 따라 문화행사 별 리스크 예측
 - 취소 리스크 별 보상범위 책정
 - 소비자를 대신에 보험사가 리스크를 담보, 취소시 발생하는 손해 보상



프로젝트 내용: 자유 주제 (1)

- 문제 제안 가이드라인
 - 기상데이터와 기타 공공데이터를 함께 활용하여 의미있는 문제를 제시하고 분석결과로 해결책 제시
 - 예시: 날씨에 따른 질환 위험도 예측, 날씨에 따른 스포츠 경기력 변화
- 프로젝트 진행 가이드라인
 - 수업과 실습에서 배운 데이터 마이닝 방법 사용한 분석
 - 데이터로부터 풀고자 하는 문제에 맞는 특징 추출, 모델링
 - 구축한 모델과 타 공공데이터를 활용한 가치 창출
- 기상청 데이터와 함께 사용할 타 공공 데이터셋 링크:
http://big.kma.go.kr/2016contest/sub_04.do
- 데이터마이닝 기법을 이용한 실전 문제 해결
- 결과 분석: 결과의 타당성 및 실용성 검증

프로젝트 내용: 자유 주제 (2)


- 기상기후 빅데이터 분석 플랫폼: 기상 데이터 제공
 - <http://big.kma.go.kr>
 - 회원가입, 공인인증서 등록, 데이터 사용권한 신청 필요
- 융합시범서비스 탭에서 특수 기상 데이터 제공

| 도메인 | 문제 |
|--------|---------------|
| 날씨와 농업 | 주산지 기상 예측 |
| | 농작물 생산성 예측 |
| 날씨와 관광 | 관광코스 기상 예측 |
| | 맞춤형 관광기후지수 추천 |
| 날씨와 수산 | 살오징어 어획량 예측 |
| 날씨와 교통 | 강수시 교통사고 예측 |
| | 교통사고 이력 분석 |

프로젝트 내용: 지정 주제

- 문제 정의
 - 관측된 기상 데이터를 사용한 집중 호우 예측
 - 전국 95개소의 종관기상관측장비(ASOS)와 493개소의 방재기상 관측장비(AWS)로부터 관측된 호우 피해 데이터 기반
 - 3시간 단위로 2006년 이후 측정된 시계열데이터의 모델링
- 호우 피해 데이터
 - 기상청 기상기후 빅데이터 분석플랫폼
 - <http://big.kma.go.kr> 홈페이지 가입 후 공인인증서 등록
융합시범서비스 >> 호우피해 예측 탭에서 다운로드
 - 5M 개의 데이터 인스턴스

프로젝트 일정

- 프로젝트/데이터 수집 설명 : 4/5
- 조 구성 마감: 4/12
- 제안서 제출 마감: 5/1
 - 문제 정의 및 해결 방안 제안
 - YSCEC 프로젝트 그룹게시판에 일정 및 조원 역할 명시
- 중간 보고서 제출 마감: 5/15
 - 진행사항 보고
 - 제안하는 방법의 아키텍처 
 - 향후 계획
- 최종 보고서 제출 마감: 5/29
 - 정의한 문제
 - 제안하는 방법과 아키텍처
 - 분석 결과와 의의