



# “국민건강 알람서비스 고도화” 완료보고



2015. 12

# '국민건강 알람서비스' 고도화 사업 완료보고회

## ● 사업 개요

○ 사 업 명 '국민건강 알람서비스 고도화 사업'

○ 사업 기간 2015. 6. 19. ~ 12. 18.(6개월)

○ 사업 비용 2억7천3백만원(연구개발비/전산개발비)

○ 사 업 자 (주)다음소프트

○ 주요 내용

-질병 예측의 정확도를 높이기 위한 새로운 예측모형 개발·적용

- 기상 및 환경 자료 등 질병발생과 상관이 있는 변수 추가
- 비정형 데이터의 수집 채널 다양화 : 인터넷 검색어, 블로그, 뉴스 등

-누구나 이해하기 쉬운 사용자 친화적인 UI 구축

## ● 추진 경과

- '15. 7. 8. 착수보고
- '15. 7. 14. 협업 관련기관 관계자 워크숍((주) 다음소프트)
- '15. 8. 11. 식중독 예보관련 업무협의(식약처)
- '15. 9. 9. 질병 예측모형 개발 중간 점검(충북대)
- '15. 9. 21. 식중독 예측모형 개발방향 협의(기상청)
- '15. 10. 20. 중간보고
- '15. 11. 25. 식중독 예측모형 개발방향 협의(개발 점검 워크숍)  
... 진료건수 기반 모형 개발
- '15. 12. 23. 감기/눈병/피부염/천식 질병 서비스 오픈  
... 식중독 예측 모델 개발 및 적용 완료

## ● 향후 계획

- 식중독 서비스 시범 운영 ( ~ '16. 2월 )
- 식중독 서비스 오픈 (3월) ... 공단, 식약처, 기상청 공동 서비스 제공

# CONTENTS

I. 사업개요

II. 사업추진 내역

# I. 사업개요

1. 사업 목표

2. 사업 주요 진행 방향

## 1. 사업 목표

예측 모델의 정확도를 향상시키고, 사용자 화면을 개선하여, 정확하고 친절한 국민건강 알람서비스를 구축

국민건강 증진 및 사회적 편익 극대화

### 정확하고 친절한 알람서비스 구축

유관기관  
협업

국민건강보험공단  
식품의약품안전처  
기상청  
국립환경과학원

정형/비정형  
데이터 보강

기상데이터  
환경데이터  
SNS데이터  
(블로그/뉴스/검색 트렌드)

예측 정확도  
향상

시계열 예측모형  
(ARIMA)  
↓  
음이항 회귀모형

사용자 화면  
개선

이용 가이드 제공  
사용자 친화적 UI / UX

## 2. 사업 추진 경과

구분	현행 서비스	고도화
알람 대상 질병	4개 (인플루엔자/눈병/식중독/피부염)	5개 (감기/눈병/식중독/피부염/천식) ※ '인플루엔자' → '감기' 변경, '천식' 추가
데이터 수집	정형 : 국민건강보험 진료 데이터 비정형 : 트위터 데이터	정형 : 진료 데이터, 식약처, 기상청, 환경부 데이터 비정형 : 네이버 검색 트렌드 뉴스, 블로그, 트위터
질병 예측 모델	ARIMAX 시계열 모델	음이항 회귀모형
위험도 모델링	질병 모두 동일한 위험도 구간	질병 특성 고려 '질병 별' 위험도 구간 설정
서비스 시각화	예측 결과 위주 학술적 서비스 제공	사용자 편의를 고려, 동선을 줄인 직관적 서비스 구성 예측 정확도 모니터링 등 편의성을 강화한 관리자 페이지 추가

## II. 사업 추진 내역

1. 알람 대상 질병 확대
2. 데이터 수집
3. 질병 예측 모델
4. 위험도 모델링
5. 서비스 시각화



## 1. 알람 대상 질병 ( 1/2 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

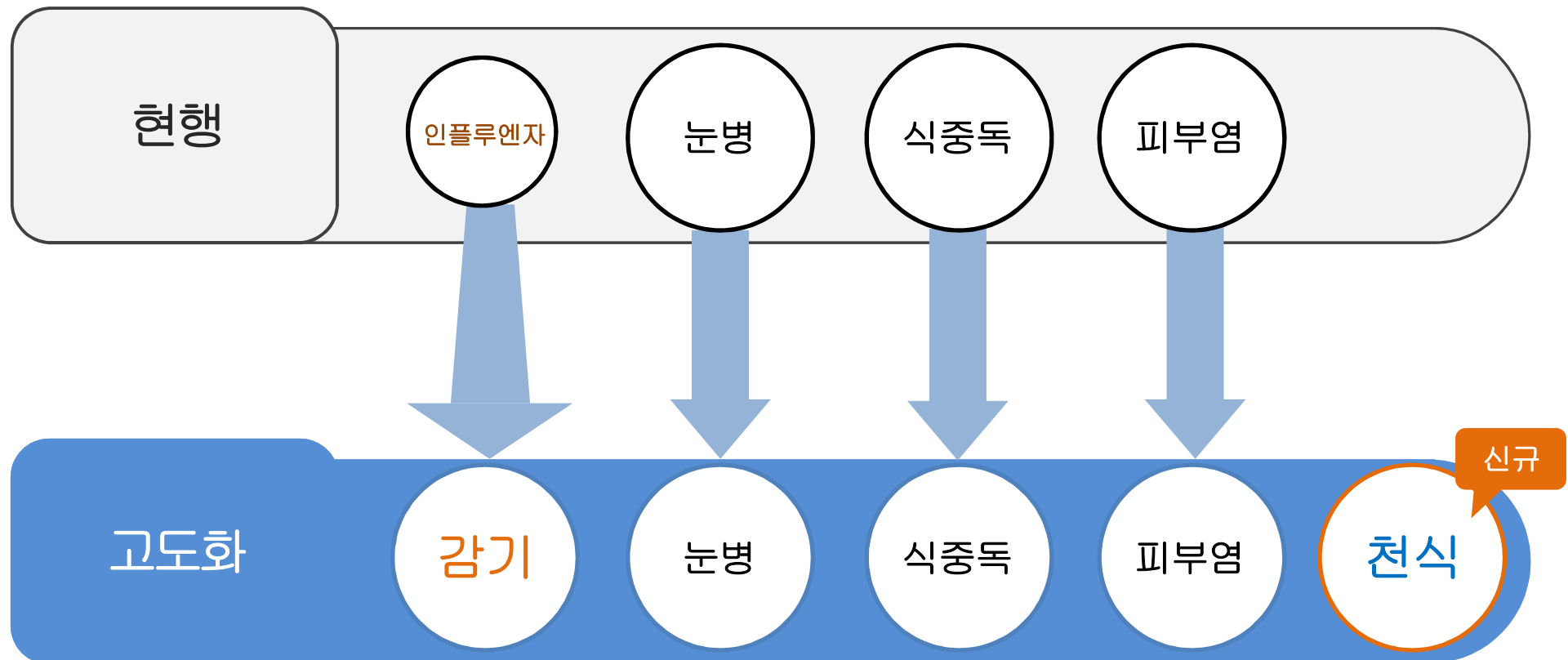
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

인플루엔자에서 감기로 질병 범위 확대, 천식이 추가되어 총 5개 질병 알람 서비스



## 1. 알람 대상 질병 ( 2/2 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

현행 서비스에서는 질병 모두 현시점 기준 알람서비스, 고도화 서비스에서는 감기, 식중독, 피부염의 경우 +2일 예보 서비스 제공

### 현행

알람

인플루엔자, 눈병, 식중독, 피부염

모두 현 시점 (오늘) **알람**

### 고도화

알람

눈병, 천식

현 시점의 상황을 **알람**

예보

감기, 식중독, 피부염

현 시점으로부터 +2일 **예측**

## 2. 데이터 수집 - 개요

질병 발생 예측 모델 정확도 개선을 위하여 정형 데이터는 식약처, 기상청, 국립환경과학원 등의 데이터, 비정형 데이터는 블로그, 뉴스, 검색트렌드(네이버)를 보강



데이터 종류	수집 채널	수집 방법	수집 간격
정형	국민건강보험공단	관리도구 FTP	일별
	식약처	관리도구	수시
	기상청	API	시간별, 일별
	국립환경과학원	API	시간별, 일별
비정형	트위터	API	일별
	블로그	API	일별
	뉴스	API	일별
	검색 트렌드	API	주별

## 2. 데이터 수집 - 프로세스

## 알람 대상 질병 확대

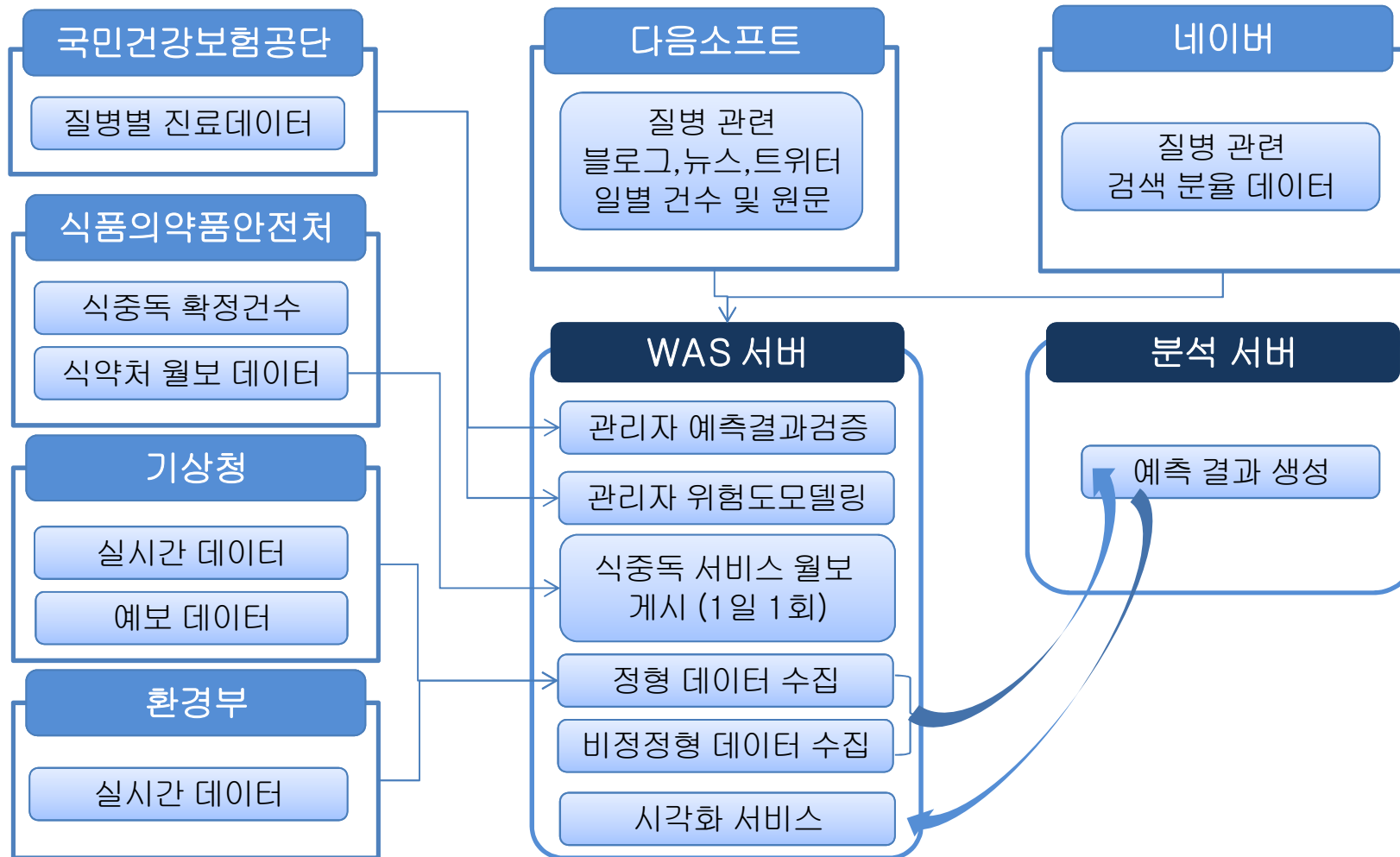
## ▶ 데이터 수집

## ▶ 질병 예측 모델

## 위험도 모델링

## 서비스 시각화

유관기관의 정형 데이터, 민간의 다양한 비정형 데이터를 일별, 시간별로 수집하여  
예측 결과 생성



## 2. 데이터 수집 - 비정형 데이터 지식분류체계 구축

비정형 데이터 수집을 위하여, 질병별 지식 분류 체계를 구축.  
질병명 지식 분류 체계와 질병별 속성을 기준으로 한 지식 분류체계로 구성됨

### 질병명

대표 질병명

질병 동의어2

질병 동의어3

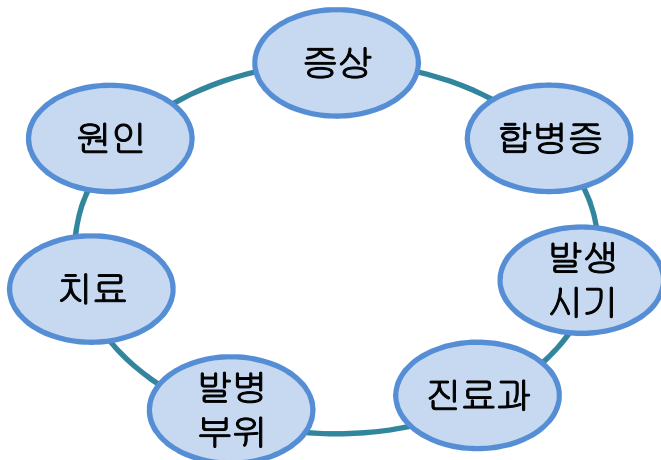
질병 동의어4

.....

### 질병명 동의어 처리 예시

대분류	중분류	소분류	대표어	동의어
질병명	감염성 질병	감기	감기	상기도감염
			신종 플루	신종 인플루엔자

### 공동 추출 정보



### 지식 분류 체계 어휘 매핑 예시

질병명	속성	대표어	동의어
감기	증상	기침나다	기침하다 기침심하다 기침계속하다
		인후통	목아프다, 편도붓다 목따갑다,편도붓다
		열나다	열오르다,발열,고열

## 2. 데이터 수집 - 비정형 데이터 수집 결과

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

### TM 2(Daumsoft) API를 통한 비정형 데이터 수집 결과

#### 블로그 / 트위터 수집 결과 요약

수집주기 : 일별

(기간 : 2011.01 ~ 현재, 단위: 천건)

구분	감기		눈병		식중독		피부염		천식	
	블로그	트위터	블로그	트위터	블로그	트위터	블로그	트위터	블로그	트위터
제외 건수	1,127	2,186	100	62	217	91	473	73	177	60
최종 건수	2,483	12,781	65	360	287	1,163	377	231	98	93

※ 제외건수 : 광고, 욕설, 스팸 등으로 인하여 삭제된 건수

#### 뉴스 수집 결과 요약

수집주기 : 일별

(기간 : 2010.07 ~ 현재, 단위: 건)

구분	감기	눈병	식중독	피부염	천식
뉴스 버즈추이	112,128	5,073	21,109	16,553	12,220

## 2. 데이터 수집 – 정형 데이터 수집 결과

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

Open API / API 를 통한 정형데이터 (식약처, 기상청, 환경부 데이터) 수집 결과

### 기상청 데이터 수집 결과

수집주기 : 매시간, 일별

( 2003년 ~ 현재 )

날짜	지역코드	기온	최고기온	최저기온	습도	최소습도	일강수량
2008.01.01	11	-6.11	-2.83	-8.43	50.08	35.34	-
2008.01.02	11	-3.78	0.96	-8.41	53.84	30.23	-
2008.01.03	11	0.67	6.75	-3.73	56.52	31.86	-

...

### 국립환경과학원 데이터 수집 결과

수집주기 : 매시간

( 2003년 ~ 현재 )

날짜	지역코드	이산화황	미세먼지	이산화질소
2008.01.01	11	0.005	34.55	0.018
2008.01.02	11	0.006	34.35	0.037
2008.01.03	11	0.010	52.09	0.052

...

### 3. 질병 예측 모델 - 개요

정형데이터와 비정형 데이터를 이용해 질병의 진료건수를 예측하는 모형으로 적합한 음이항 회귀 모형을 선택

현행

시계열 예측 모형 - ARIMA



- 진료 건수 , 소셜 (트위터) 기반 예측 모델
- 업무 특성상 45~90일 지연 취합되는 진료건수로 인해 **예측력이 저하**

고도화  
알람서비스

감기

눈병

식중독

피부염

천식

음이항 회귀 모형



- 예측하고자 하는 변수(진료건수)에 유의하게 영향을 미치는 변수들(기상변수, 환경변수, 소셜데이터)을 결합한 예측 모형
- 지연없이 취합되어 지는 데이터를 바로 예측 모델에 사용



### 3. 질병 예측 모델 - 식중독 세부사항

식중독 서비스의 경우 식약처의 확진건수 정보, 기상, 환경, 소셜 데이터를 융합한 예측모델 기반으로 공단, 식약처, 기상청 3개 기관이 동일하게 식중독 지수 서비스

#### 건강보험공단 / 식약처 / 기상청 / 국립환경과학원 4개 기관 협업

식약처, 기상청,  
국립환경과학원

식중독 확진건수, 기상, 환경 데이터 제공

건강보험공단

유관기관 정형데이터 + 소셜 데이터 융합 예측 모형 개발

식약처  
기상청

공단에서 개발한 예측 모델을 활용, 식중독 지수 자체 생성 및 공유



\* 식중독의 경우 약 2개월 간의 시범 운영 기간을 거쳐 3월경 오픈할 예정

### 3. 질병 예측 모델 - 세부사항 (1/2)

전국 예측모델과 **지역적 유사성을 가지고 있는 7개 권역으로 구분**

※ 7개 권역 : 서울/경기/인천 , 강원, 충남/충북/대전, 경북/대구, 경남/울산/부산, 전북/전남/광주, 제주

7개 권역별로 동일한 기상변수 및 환경변수를 사용하는 예측 모델 개발

질병	기상청 (7개 변수)							국립환경과학원 (3개 변수)			SNS (3개 변수)		
	평균 기온	최고 기온	최저 기온	일교차	습도	강수량	풍속	미세 먼지	이산 화황	오존	트위터	뉴스	네이버
감기			○	○	○						○	○	
눈병		○				○		○				○	○
피부염		○										○	
천식			○	○	○			○	○	○	○	○	○
식중독	○		○	○	○	○	○	○			○	○	

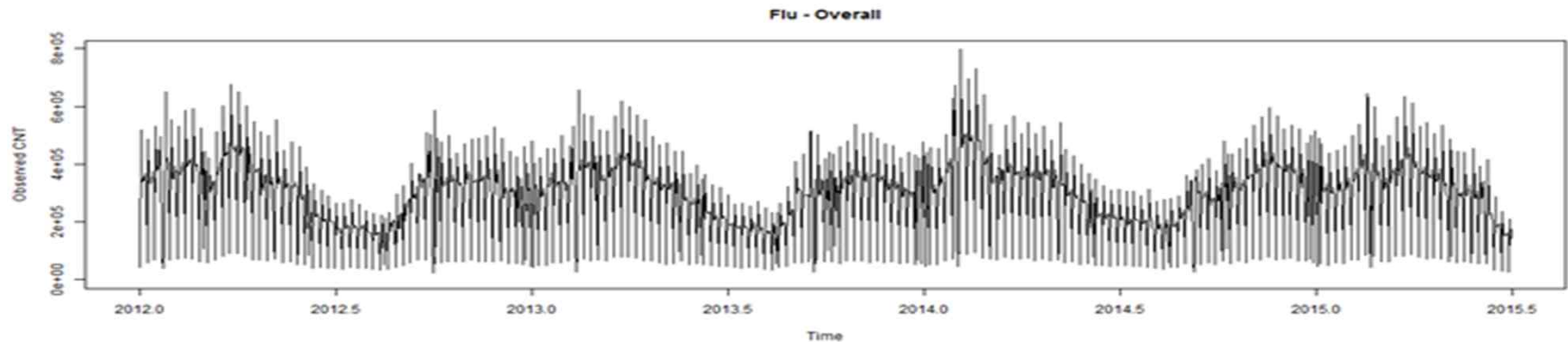
### 3. 질병 예측 모델 - 세부사항 (2/2)

예측 모델 구동 결과 실제값과 예측값 사이에 0.9 이상의 높은 상관도를 나타냄

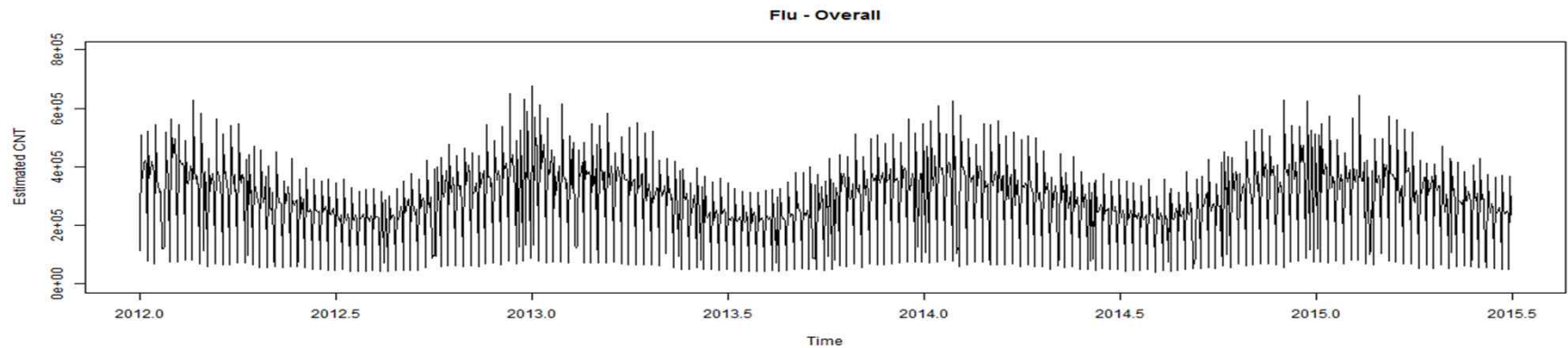
전국기준

구분	감기	눈병	식중독	피부염	천식
상관도	0.912	0.97	0.9157	0.971	0.933

(감기)  
실제 값



(감기)  
예측 값

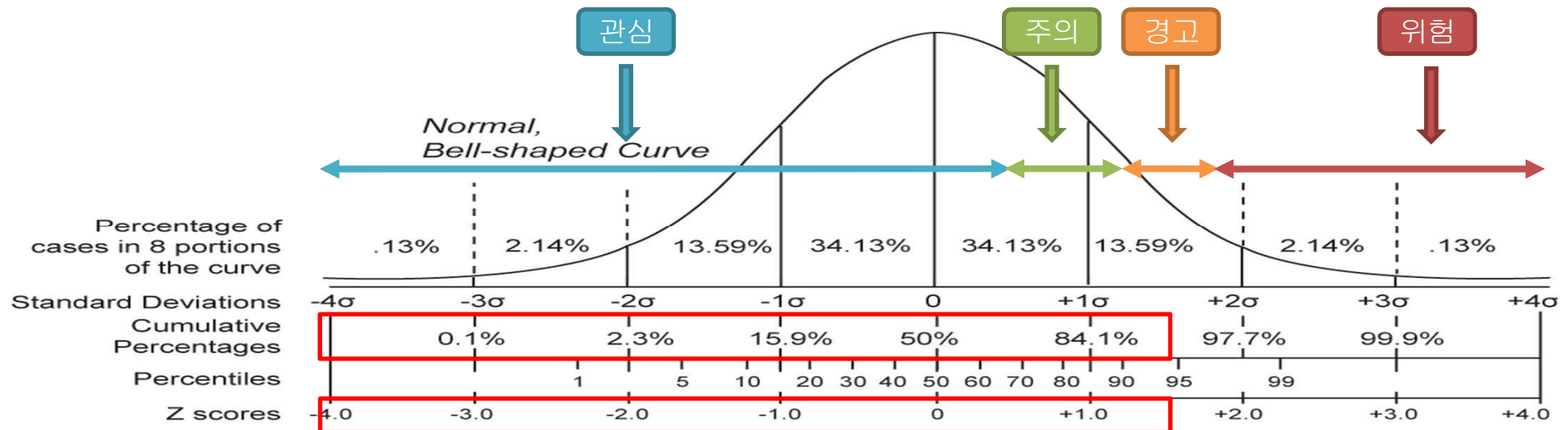


## 4. 위험도 모델링 – 기존의 위험도 모델링

기존에는 질병 별로 동일하게 적용된 Z-Score 위험도 구간을 설정

$$Z = \frac{X - \mu}{\sigma} \quad \text{where } \mu = \text{평균}, \sigma = \text{표준편차}$$

단계	구간	구간 확률	누적 확률
관심	$-\infty \leq Z \leq 0.50$	69.15%	69.15%
주의	$0.50 < Z \leq 1.28$	20.82%	89.97%
경고	$1.28 < Z \leq 1.89$	7.09%	97.06%
위험	$1.89 < Z \leq \infty$	2.94%	100.00%



## 4. 위험도 모델링 - 질병별 위험도 구간 세분화

### 질병별 특성을 고려한 위험도 구간 설정

#### ■ 감기

단계	구간	구간 확률	누적 확률
관심	$-\infty \leq Z \leq -0.14$	40.00%	40.00%
주의	$-0.14 < Z \leq 0.60$	30.00%	70.00%
경고	$0.60 < Z \leq 1.13$	20.00%	90.00%
위험	$1.13 < Z \leq \infty$	10.00%	100%

#### ■ 눈병

단계	구간	구간 확률	누적 확률
관심	$-\infty \leq Z \leq -0.22$	40.00%	40.00%
주의	$-0.22 < Z \leq 0.46$	30.00%	70.00%
경고	$0.46 < Z \leq 1.31$	20.00%	90.00%
위험	$1.31 < Z \leq \infty$	10.00%	100%

#### ■ 식중독

단계	구간	구간 확률	누적 확률
관심	$0 \leq P \leq 0.40$	40.00%	40.00%
주의	$0.40 < P \leq 0.70$	30.00%	70.00%
경고	$0.70 < P \leq 0.90$	20.00%	90.00%
위험	$0.90 < P \leq 1$	10.00%	100%

#### ■ 천식

단계	구간	구간 확률	누적 확률
관심	$0 \leq Z \leq -0.09$	25.00%	25.00%
주의	$-0.09 < Z \leq 0.72$	25.00%	25.00%
경고	$0.72 < Z \leq 1.17$	30.00%	80.00%
위험	$1.17 < Z \leq \infty$	20.00%	100%

#### ■ 피부염

단계	구간	구간 확률	누적 확률
관심	$-\infty \leq Z \leq -0.51$	40.00%	40.00%
주의	$-0.51 < Z \leq 0.84$	30.00%	70.00%
경고	$0.84 < Z \leq 1.34$	20.00%	90.00%
위험	$1.34 < Z \leq \infty$	10.00%	100%

## 5. 서비스 시각화 - 기존 사용자 화면

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

질병에 대한 지역별 위험도 및 위험도 단계별 행동요령의 인식성이 떨어짐.  
사실 전달 위주의 화면 구성





## 5. 서비스 시각화 - 사용자 화면 ( 1/5 )

국민건강 알람서비스 고도화 사업

### II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

사용자 친화적인 서비스를 위하여 가이드 제공, 직관적인 UI 로 사용자 화면 구성



## 5. 서비스 시각화 - 사용자 화면 ( 2/5 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

식중독의 경우, 식중독 지수 및 식약처에서 제공한 월별 식중독 안내 서비스





## 5. 서비스 시각화 - 사용자 화면 ( 3/5 )

서비스 이용 방법에 대한 상세한 가이드 및 위험도 단계에 대한 상세 설명을 제공함으로써 대국민 이해도 증가 도모

**국민건강 알람서비스 | 이용가이드**

국민건강 알람서비스 1 감기 2 눈병 3 식중독 4 천식 5 피부염 10 서비스개요

2015.12.01

오늘 내일 모레 4

전국 감기는 관심 단계입니다.

한기를 자주 시켜 깨끗한 환경을 유지하고, 외출 후에는 반드시 손을 씻는 등 평소 손 씻기를 생활화합니다.

제주

**국민건강 알람서비스 | 위험도가이드**

국민건강 알람서비스 위험도가이드 인쇄

감기 눈병 식중독 천식 피부염

**○ 감기 위험도란?**

국민건강보험공단이 보유한 국민건강정보DB와 기상청에서 제공하는 기상기상관측정보, 민간의 소셜미디어 정보를 통해 지역별 진료정보를 예측하여, 지역별로 감기에 대한 발병가능성 및 유행정도 등의 예측정보를 관심, 주의, 경고, 위험 4단계로 알려주는 척도입니다.

**○ 감기 예측진료건수**

질병발생에 영향을 미치는 기상기상관측정보인 최저기온, 일교차와 소셜정보인 트위터, 뉴스 발생건수를 기반으로 국민건강보험공단 진료데이터의 요일효과를 적용하여 예측되는 감기와 관련된 질병의 예측진료건수입니다.

- 요일효과 : 주말엔 발생하지 않는 진료 건수를 고려하여 7일 이동평균을 사용하여 요일별 진료 건수를 평준화 시킨 것

**○ 감기 진료발생건수 예측모델**

$$y = \exp(1.169e+01 - 1.527e-02 \cdot \text{최저기온} + 1.738e-02 \cdot \text{일교차} + \text{요일효과} + 5.131e-06 \cdot \text{트위터} + 2.793e-04 \cdot \text{뉴스})$$

- 위 모델은 전국을 기준으로 한 모델이며 시도 및 시군구 단위를 기준으로 한 예측모델은 포함 데이터 및 계수에서 차이가 날 수 있습니다.
- y = 예측진료건수
- e : 십의 차승을 의미합니다. (예 :  $1.0E+03 = 1000$  /  $1.0E-03 = 0.01$ )

**○ 감기 위험도 예측**

예측된 질병 발생 건수를 과거 2년 평균 질병 발생 건수와 비교하여, 평균으로부터 벗어나 있는 정도를 기준으로 해당값의 분포도에 따라 위험도를 관심, 주의, 경고, 위험 4단계로 구분합니다.

## 5. 서비스 시각화 - 사용자 화면 ( 4/5 )

국민건강 알람서비스 고도화 사업

### II. 사업 추진 내역



알람 대상 질병 확대

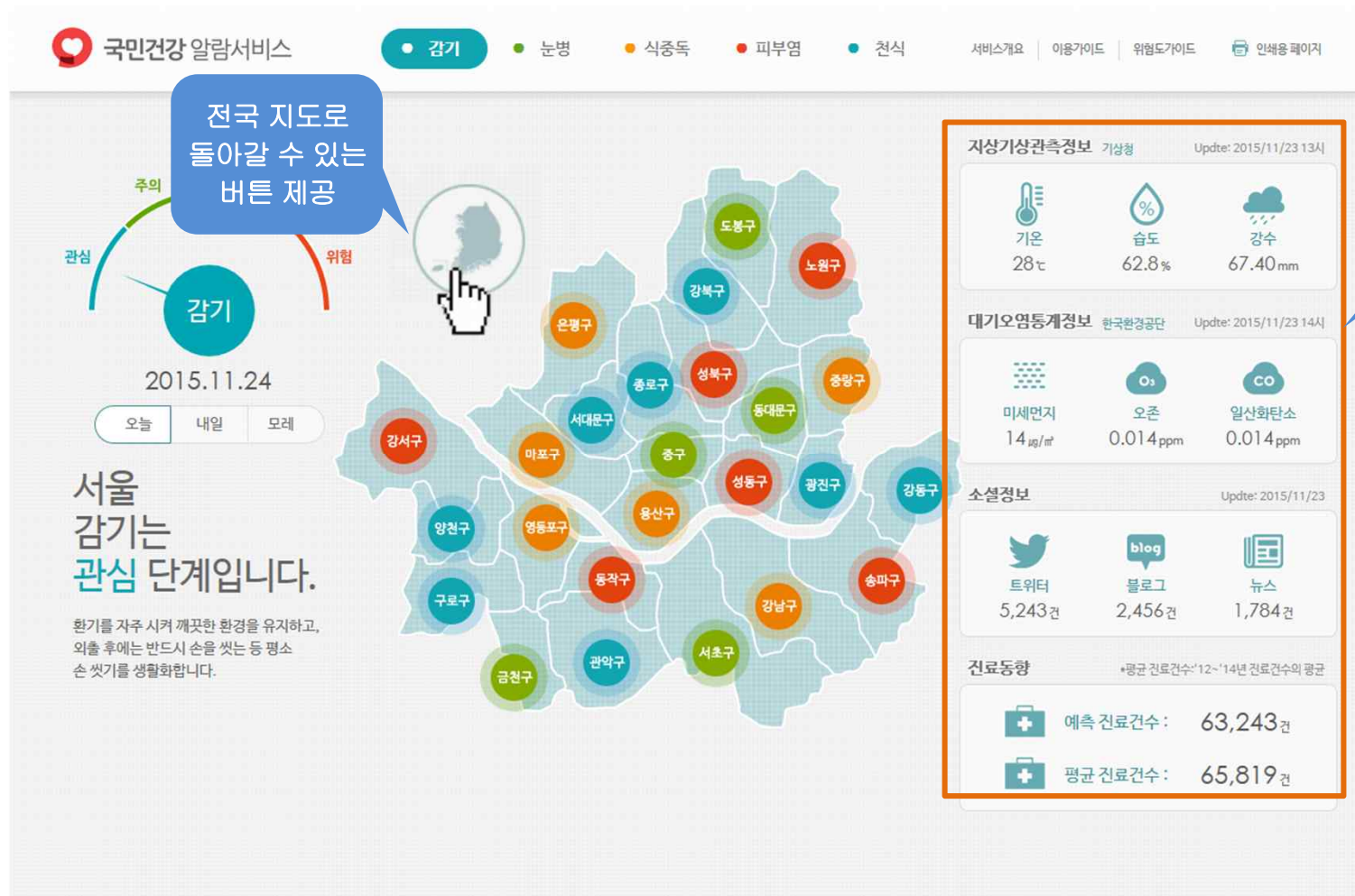
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

시도 확대 지역을 제공하여, 시군구 단위까지 상세 조회 가능



## 5. 서비스 시각화 - 사용자 화면 ( 5/5 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

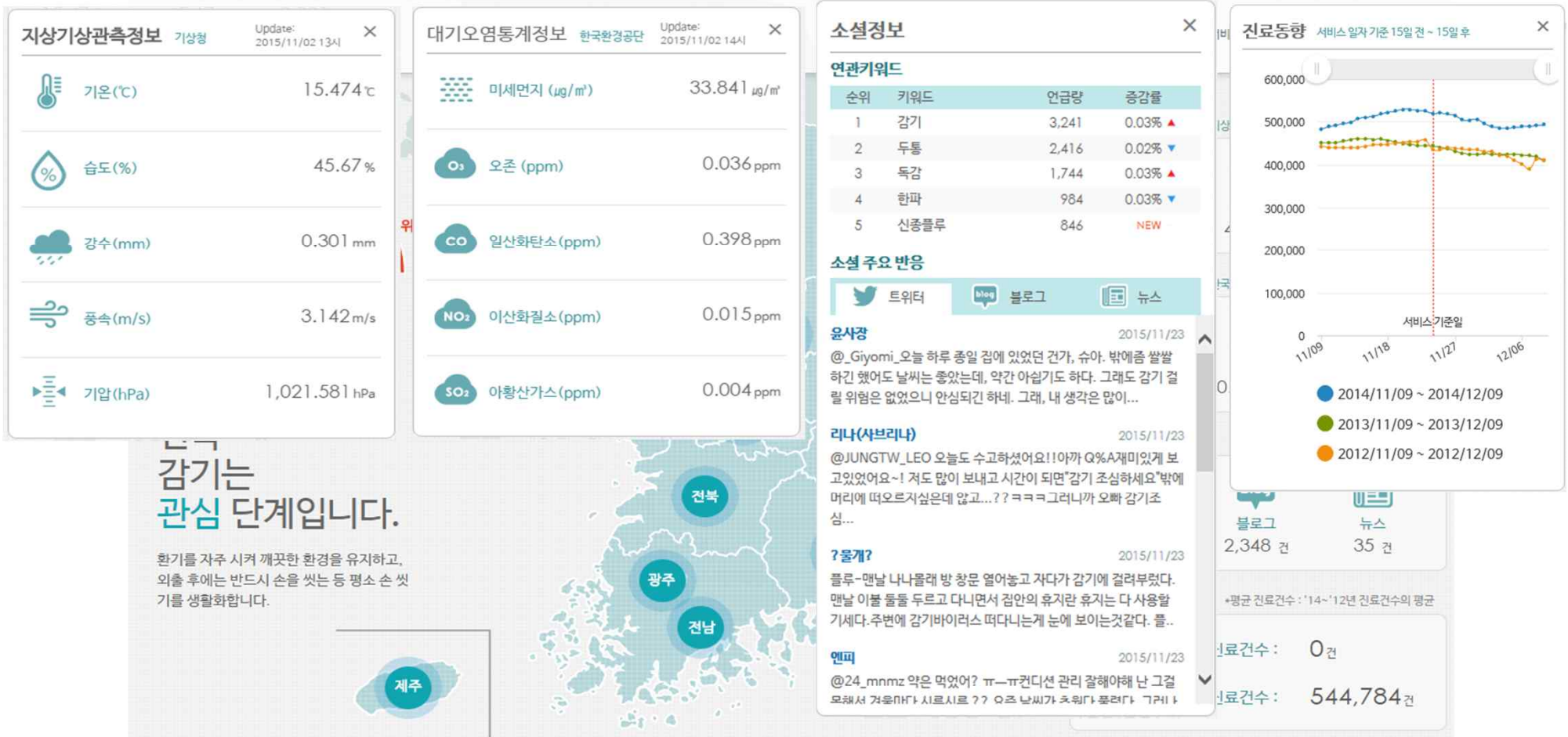
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

각 기상정보, 환경정보, 소셜정보, 진료동향 등은 마우스 클릭 시 아래와 같이 상세 정보를 제공



## 5. 서비스 시각화 - 관리자 화면 ( 1/4 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

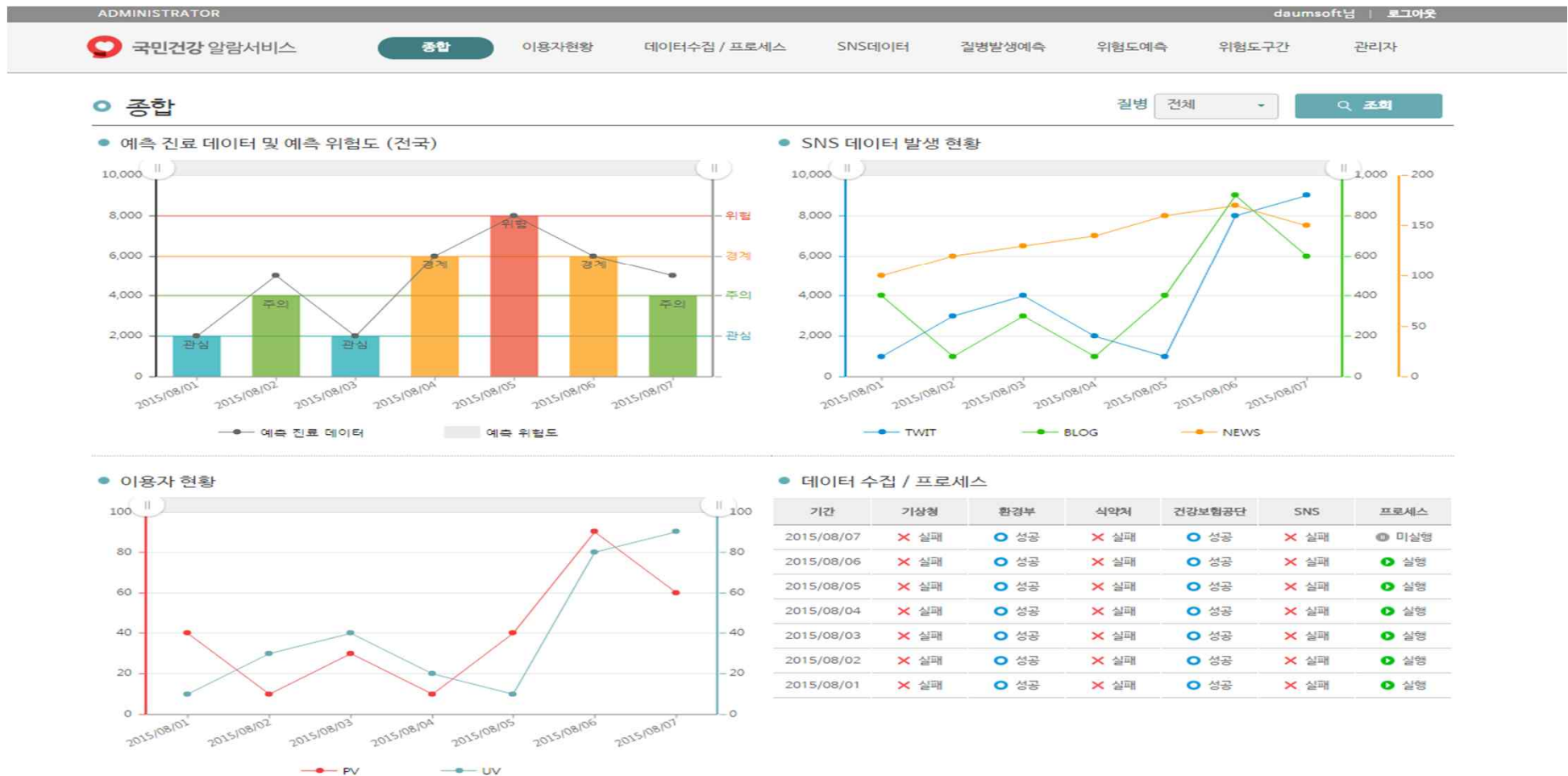
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

종합 화면을 통해 최근 일주일간의 위험도 예측 현황, 데이터 수집 현황, 이용자 현황 모니터링 기능 제공





## 5. 서비스 시각화 – 관리자 화면 ( 2/4 )

‘데이터수집/프로세스’ 페이지를 통하여 데이터 수집 현황 및 예측프로세스 실행 현황 모니터링 기능 제공. ‘업로드’, ‘다운로드’ 기능을 통해 데이터 관리 가능

ADMINISTRATOR daumsoft님 | 로그아웃

국민건강 알람서비스 종합 이용자현황 **데이터수집 / 프로세스** SNS데이터 질병발생예측 위험도예측 위험도구간 관리자

데이터수집 / 프로세스 기간 2015/08/10 ~ 2015/09/10 **조회**

기간	기상청	환경부	식약처	건강보험공단	SNS	프로세스
2015/08/07	× 실패	○ 성공	× 실패	○ 성공	× 실패	미실행
2015/08/06	× 실패	○ 성공	× 실패	○ 성공	× 실패	실행
2015/08/05	× 실패	○ 성공	× 실패	○ 성공	× 실패	실행
2015/08/04	× 실패	○ 성공	× 실패	○ 성공	× 실패	실행
2015/08/03	× 실패					실행
2015/08/02	× 실패					실행
2015/08/01	× 실패					실행
2015/08/07	× 실패					미실행
2015/08/06	× 실패					실행
2015/08/05	× 실패					실행
2015/08/04	× 실패					실행
2015/08/03	× 실패					실행
2015/08/02	× 실패	○ 성공	× 실패	○ 성공	× 실패	실행
2015/08/01	× 실패	○ 성공	× 실패	○ 성공	× 실패	실행

**업로드**

날짜 2015/09/10

종류 기상청

파일

**다운로드**

종류 기상청

기간 2015/09/10 ~ 2015/09/10

## 5. 서비스 시각화 – 관리자 화면 ( 3/4 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

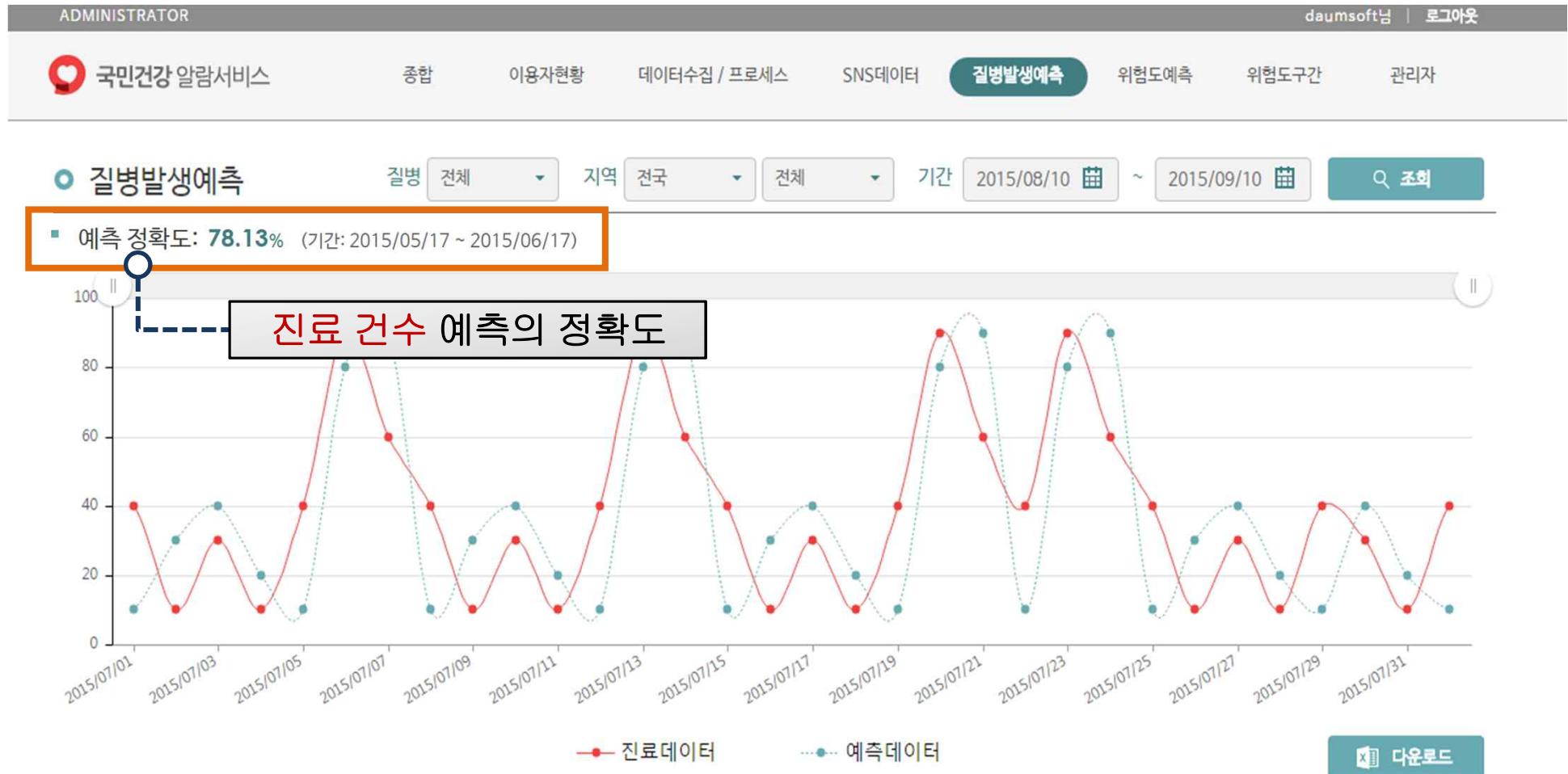
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

조회 기간의 실제 진료 건수와 예측 진료 건수 간의 비교 및 추이 조회



## 5. 서비스 시각화 - 관리자 화면 ( 4/4 )

국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대

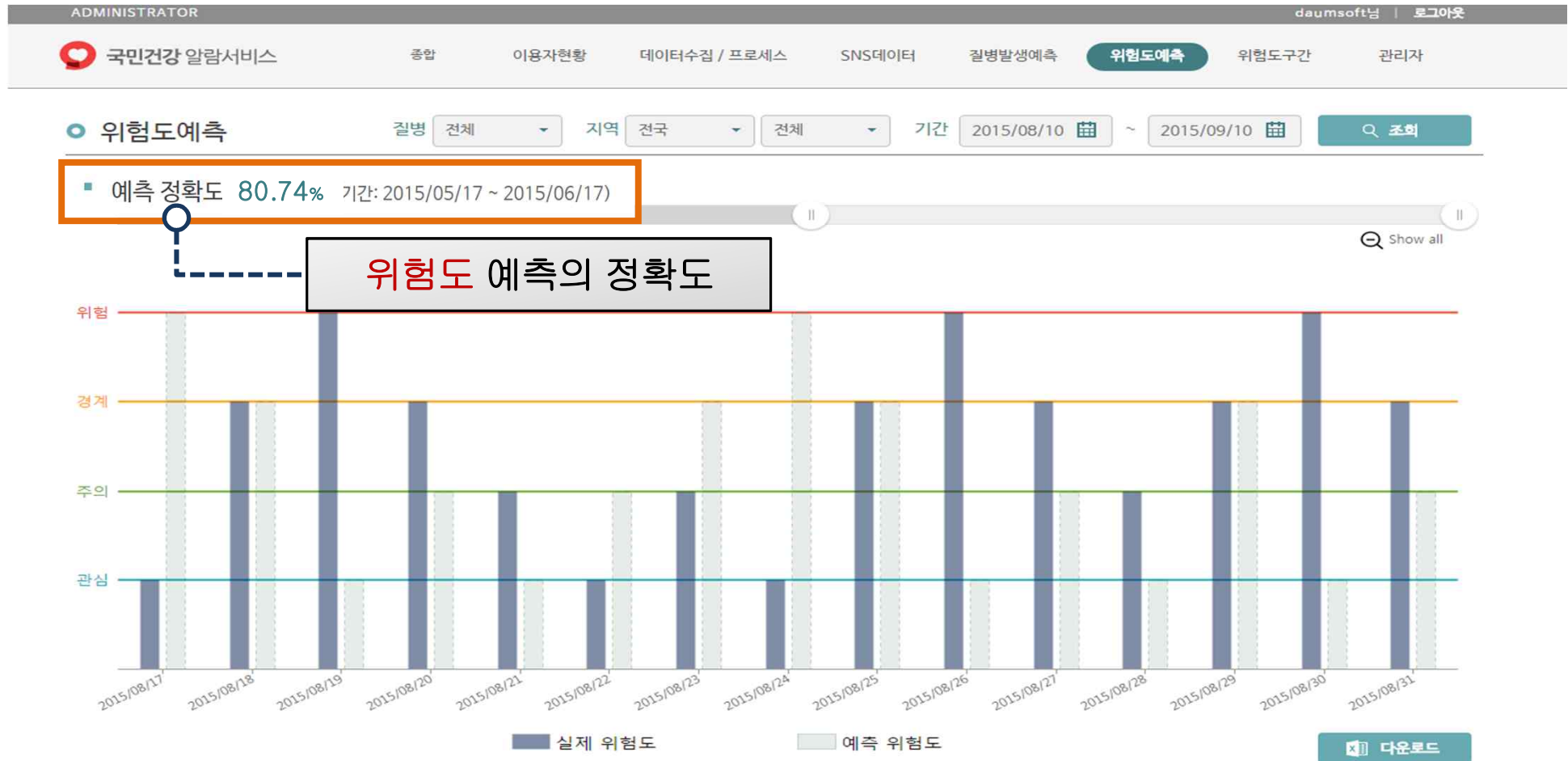
데이터 수집

질병 예측 모델

위험도 모델링

서비스 시각화

조회 기간의 실제 위험도와 예측 위험도 간의 비교 및 추이 조회





시 연

*Daumsoft*  
MINING MINDS





감사합니다.

*Q & A*

***Daumsoft***  
MINING MINDS