

# 사회보장 빅데이터를 활용한 복지사각지대 발굴 사례

2017. 10. 26.

오미애 한국보건사회연구원 빅데이터연구팀장

# 목차

## 1. 연구배경

---

## 2. 개념 정의

---

## 3. 복지사각지대 발굴을 위한 연계 정보

---

## 4. 복지사각지대 발굴 프로세스

---

## 5. 복지사각지대 발굴 예측 모형

---

## 6. 개선 사항 및 향후 방향

# 1. 연구배경

---

# 1 연구배경

## 송파 세 모녀 자살 사건

위키백과, 우리 모두의 백과사전.

송파 세 모녀 자살 사건은 2014년 2월 송파구에 사는 세 모녀가 큰딸의 만성 질환과 어머니의 실직으로 인한 생활고에 시달리다가 “정말 죄송합니다”라는 메모와 함께 갖고 있던 전 재산인 현금 70만원을 집세와 공과금으로 놔두고 **번개탄**을 피워 자살한 사건이다. 세 모녀는 부양의무자 조건 때문에 **국민기초생활보장제도**의 도움을 받지 못하고 있었다. 이들은 자살하기 3년 전 관공서에 복지 지원을 타진했으나 대상 조건을 만족하지 못한다는 것을 알게 된 뒤 재신청을 하지 않고 생활해 왔다. 이는 30세 성인에 대한 추정소득이 산정되었기 때문이다.

- 2014년 12월 송파 세 모녀법으로 불리는 「국민기초생활 보장법」 및 「긴급복지 지원법」 개정안, 「사회보장급여의 이용·제공 및 수급권 발굴에 관한 법률」 제정 등 복지 사각지대 해소와 관련된 3개 법안이 국회를 통과, 2015년 7월 1일부터 시행

# 1 연구배경

## 「사회보장급여의 이용·제공 및 수급권 발굴에 관한 법률」

가. 이 법은 **사회보장 관련 정보 또는 신청능력의 부족으로 보호받지 못하는 사회보장수급권자를 발굴하여 지원**함으로써 국민의 사회보장수급권을 최대한 보장하고 사회보장급여가 공정하고 효과적으로 제공하는 것을 목적으로 함(안 제1조)

나. 수급권자의 발굴을 위하여 **사회복지법인, 국민연금공단, 보건소 또는 경찰서 등의 기관 및 단체 간의 연계 및 협력, 관련 정보의 공유** 등이 원활하게 이루어질 수 있도록 하고, 이들 정보를 사회보장 사각지대 해소를 위해 **제공될 수 있도록 함**(안 제4조 및 제7조부터 제9조)

다. 누구든지 사회적 위험에 처한 보호대상자를 발견하였을 때에는 보장기관에 신고하도록 하고, 신고된 보호대상자에 대하여 조사·상담·안내·의뢰하며, 보호대상자가 사회보장급여를 신청하는 데 필요한 서비스를 제공하도록 함(안 제9조부터 제14조)

# 1 연구배경

## 「사회보장급여의 이용·제공 및 수급권 발굴에 관한 법률」

라. 국가 및 지방자치단체는 수급권자별 사회보장급여 제공계획을 수립하고 **제공결과를 정기적으로 평가**하여 그 결과에 따라 보호계획을 변경하도록 하며, 이들 업무를 실시하기 위하여 **전담하는 직원**을 두도록 함(안 제15조)

마. 수급권자의 발굴, 의뢰, 신청, 조사, 결정, 보호계획의 수립 등의 업무에 종사하거나 종사하였던 자는 **사회보장업무 수행과 관련하여 알게 된 개인·법인 또는 단체의 정보를 관계 법령에서 정하는 바에 따라 보호하도록 함**(안 제17조)

바. 국가 및 지방자치단체의 장의 권한 또는 업무의 일부를 소속 기관의 장이나 사회복지 관련 기관 및 단체에 위탁하고, 그 업무의 처리에 필요한 인력 또는 경비를 지원할 수 있는 근거를 마련함(안 제19조 및 제20조)

# 1 연구배경

## 추진 경과

- ✓ 복지사각지대로 인한 사회적 문제가 지속적으로 제기됨에 따라 지자체 일제조사를 통한 사각지대 대상자 발굴 추진('11년~)
  - 지자체 공무원, 통. 리반장 등이 참여한 민관협력을 통해 지역사회 취약계층을 발굴하기 위한 복지사각지대 일제조사
- ✓ '15년 보건복지부 중점 업무계획으로 「사각지대 관리시스템 구축을 통한 잠재 위기가구 주기적 모니터링」 과제 제시('15.1월)
  - 다양한 공공 정보를 수집, 분석하여 위기가구를 발굴하고, 지자체를 통해 복지 지원을 연계하며, 그 지원 이력을 지속 관리
- ✓ 2015년도부터 보건복지부, 한국보건사회연구원, 사회보장정보원, 서울대학교 통계학과가 협업하여 복지사각지대 발굴 방안 연구 진행 중
  - 사회보장정보시스템을 활용한 복지사각지대 발굴방안 연구('15.1 ~ 9월)
  - 사회보장정보시스템을 활용한 복지사각지대 발굴방안 후속연구('16.3 ~ 12월)
  - 사회보장정보시스템을 활용한 복지사각지대 발굴방안 확대 연구('17~ )

## 2. 개념 정의

---

개념 I 사회보장정보시스템

개념 II 복지 사각지대

개념 III 사회보장 빅데이터



# 2 개념 정의

## 사회보장정보시스템

### 주요사업

사회보장정보시스템

지역보건의료정보시스템

사회서비스 전자바우처

보육통합정보시스템

사회복지시설정보시스템

취약계층 지원시스템

보건복지 포털

사회보장 정보화 정책 연구

개인정보 비식별 조치지원

### 사회보장정보시스템



정부 각 부처에서 분산되어 운영되고 있는 복지사업과 지원대상자의 정보를 통합관리하여 복지 업무를 처리하고 적정수급관리를 지원하는 시스템

사회보장정보시스템(행복e음)

사회보장정보시스템(범정부)

자주묻는질문



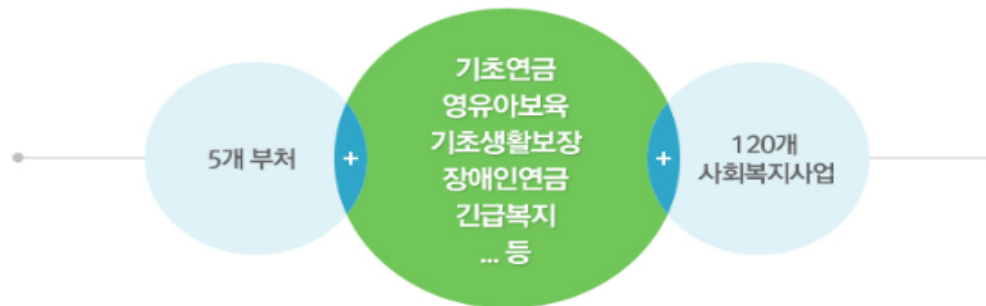
사회보장정보시스템 2014년 UN공공행정상 대상 수상

### 시스템소개

각종 사회복지 급여 및 서비스 지원 대상자의 자격과 이력에 관한 정보를 통합 관리하고, 지자체의 복지업무 처리를 지원하기 위해 기존 시·군·구별 세물행정시스템의 31개 업무 지원시스템 중 복지분야를 분리하여 개인별 가구별 DB로 중앙에 통합 구축한 정보시스템입니다.

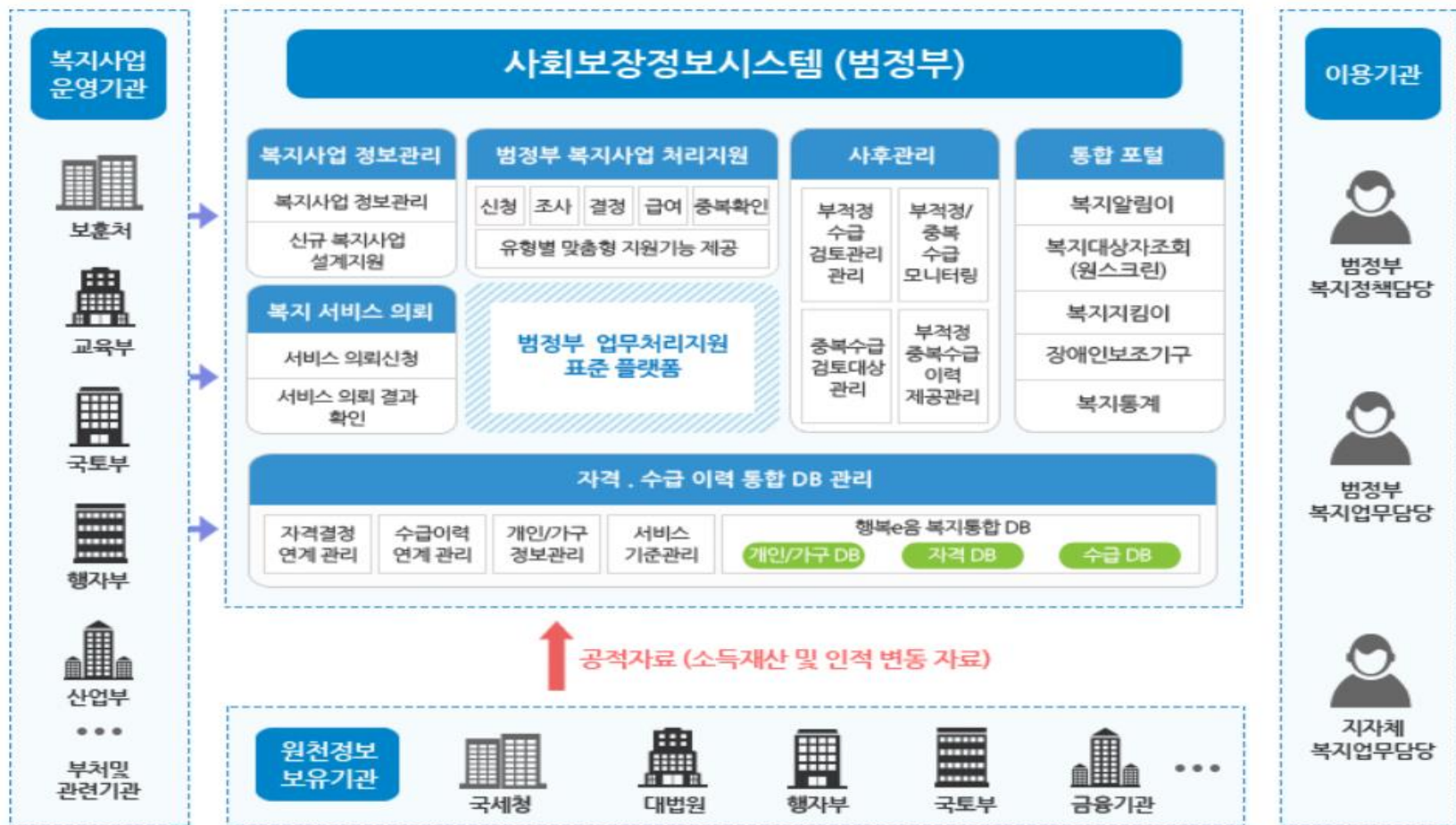
▶ 사회보장정보시스템(행복e음) 콜센터 : 1566-3232 (단축번호 2번)

### 주요서비스



# 2 개념 정의

## 사회보장정보시스템



## 2 개념 정의

### 복지 사각지대

- ✓ 사회보장정보시스템을 고려할 때의 복지 사각지대의 범위는 사회보장정보시스템에서 관리하는 복지제도의 지원 기준에 적합하지만 해당 복지 급여의 지원을 받지 않는 개인이나 가구를 의미
- ✓ 사회보장정보시스템에서 관리하는 복지제도의 지원 기준에 적합하지 않지만, 업무 담당자가 민간자원을 통해 한시적 도움을 제공할 필요가 있다고 판단되는 개인이나 가구를 의미

☞ 본 연구에서는 복지 사각지대의 분석 대상을 사회보장정보시스템에서 접근 가능한 차상위계층까지 한정

## 2 개념 정의

### 사회보장 빅데이터

- ✓ 사회보장 빅데이터의 개념은 사회보장의 정의에 따라 사회보장이 포괄하는 영역의 범위와 관련되며, “사회보장 분야에서 생산 및 구축된 행정데이터 또는 빅데이터”로 정의할 수 있음
  - 이는 사회보장기본법 제3조에 제시된 사회보장의 정의를 근거로 설정할 수 있는데, “사회보장”은 출산, 양육, 실업, 노령, 장애, 질병, 빈곤 및 사망 등의 사회적 위험으로부터 모든 국민을 보호하고 국민 삶의 질을 향상시키는 데 필요한 소득·서비스를 보장하는 사회보험, 공공부조, 사회서비스를 의미함(제3조의1)
- ✓ 사회보장기본법에 기초하여 사회보장의 개념과 범위를 정의할 수 있는 것처럼, 행정데이터 또는 빅데이터를 의미하는 정보 및 이러한 정보가 구축 및 운영되는 시스템은, 사회보장급여의 이용·제공 및 수급권자 발굴에 관한 법률(사회보장급여법) 제23조의 “사회보장정보”와 사회보장기본법 제37조에 근거한 “사회보장정보시스템(행복e음, 범정부시스템)”으로 정의된다고 할 수 있음

## 2 개념 정의

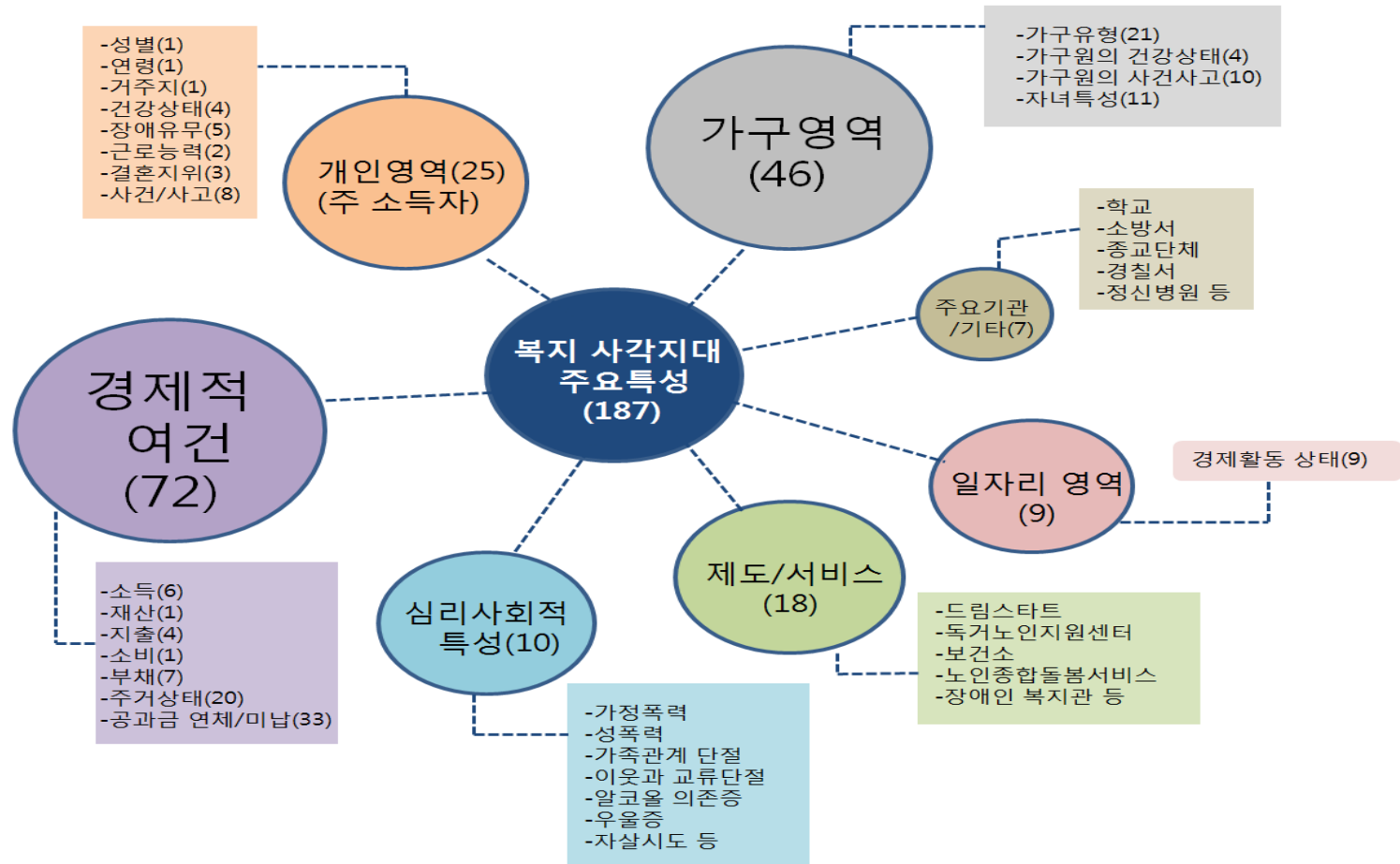
### 사회보장 빅데이터

따라서, 사회보장 빅데이터의 개념과 범위는 “사회보장 분야(사회보장기본법 제 3조)의 다양한 정책수립 및 시행과 관련하여 정책의 집행과정에서 사회보장정보 시스템(사회보장기본법 제37조)과 각종 사회보험 제도 운영을 지원하는 정보시스템을 통해 구축된 사회보장정보(사회보장급여법 제23조제1항)”로 설정할 수 있음

### 3. 복지사각지대 발굴을 위한 연계 정보

---

# 3 복지사각지대 발굴을 위한 주요 항목



# 3 복지사각지대 발굴 연계 정보

구분		대상 기관
개인 및 가구영역	범죄 피해자 정보	경찰청
	화재 피해자 정보	국민안전처
	재난 피해자 정보	
	주거위험 1(전세보증금)	국토교통부
	주거위험 2(월세정보)	
	위기학생 정보	교육부
일자리 영역	고용위험1 (개별연장급여 대상)	고용노동부
	고용위험2 (임금체불 등 사유로 실업급여 수급)	
	고용위험3 (실업급여 수급기관 과소자)	
	고용위험4 (실업급여 비대상자)	



### 3 복지사각지대 발굴 연계 정보

구분		대상 기관
경제적 여건	단전 정보	한국전력공사
	전기료 체납정보	
	단수 정보	상수도사업본부
	단가스 정보	도시가스 사업자
	건보료 체납 가구	건강보험공단
	의료비 과다 지출가구(본인부담금 상한액)	
	장기요양보험 수급상태	
	국민연금 보험료 체납	국민연금 관리공단
심리·사회적 특성	자살 고위험군	보건소, 자살예방 센터
	자살, 자해 시도자 정보	응급의료 센터
제도/서비스 등	건강 위험1 (방문건강사업 집중관리 대상)	보건소
	건강 위험2 (미숙아지원사업)	
	시설입소 및 퇴소자	복지부
	기초생활수급/긴급복지 탈락 및 중지자	

# 3 복지사각지대 발굴 연계 정보

## 외부기관

한국전력공사

건강보험공단

국민연금공단

교육부

광역상수도사업본부

국민안전처

고용노동부

응급의료센터

경찰청

국토교통부

보건소

도 상수도사업소

도시가스사업자

자살예방센터

근로복지공단

신용정보원

사회보장  
빅데이터

## 외부연계정보(23종)

- 단전, 단수, 단가스 데이터
- 건보료 체납 데이터
- 의료비 과다지출 데이터
- 국민연금 체납 데이터
- 위기 학생 데이터
- 건강 위험 데이터
- 범죄 피해 데이터
- 화재 피해 데이터
- 주거, 고용 위험 데이터
- 자살 위험 데이터
- 공공임대주택 체납 데이터
- 금융연체 데이터
- 재난데이터
- 산재요양종결 데이터

## 사회보장시스템 내부정보(33종)

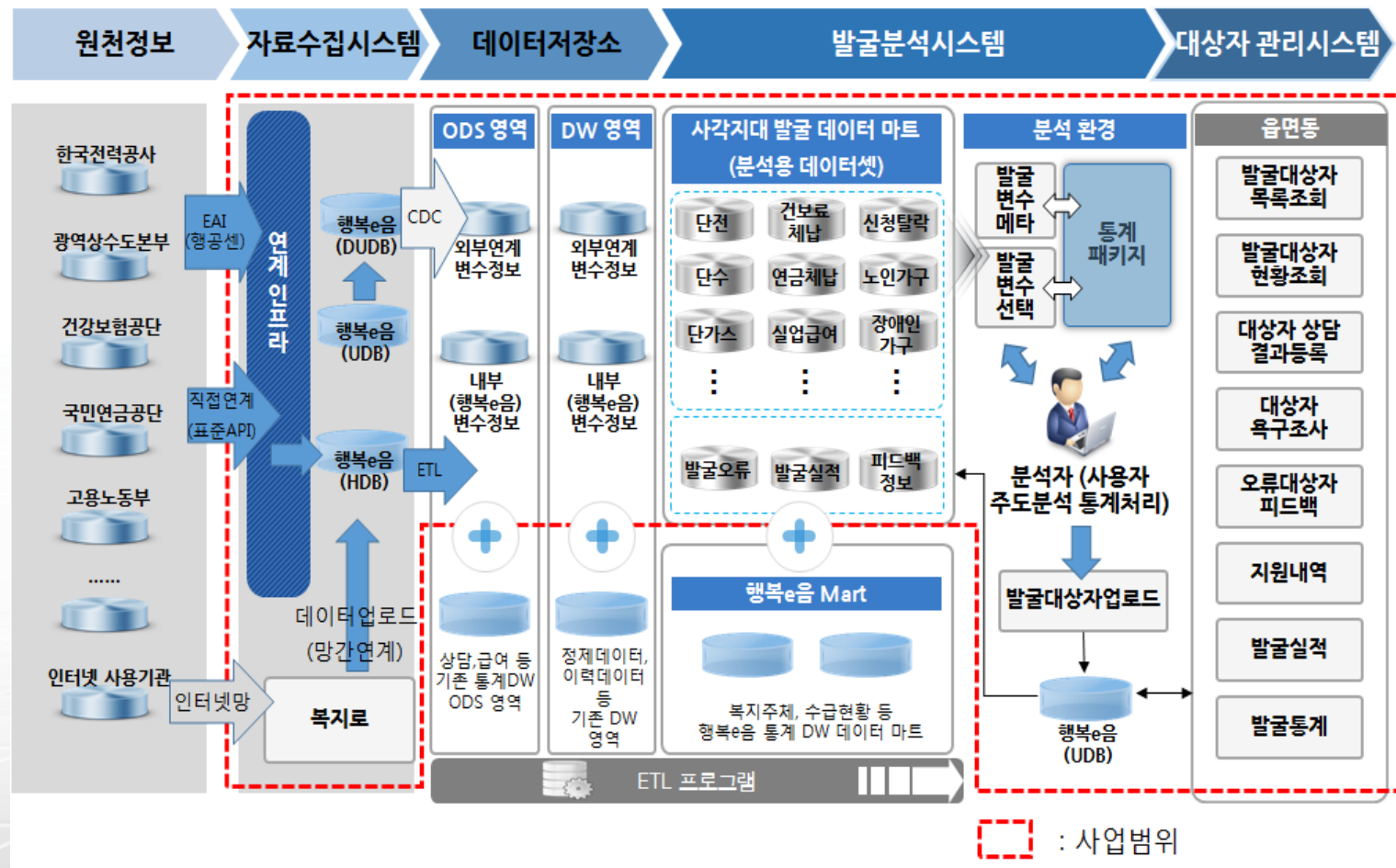
- 개인 건강상태
- 개인 장애유무
- 장애인 가구 여부
- 일자리 정보
- 소득인정액
- 가구유형 정보 등



## 4. 복지사각지대 발굴 프로세스

---

# 4 복지사각지대 발굴관리 목표시스템 개념도



# 4 복지사각지대 발굴 프로세스

## 복지 사각지대 고위험가구 예측 및 발굴 프로세스

(step1) 전체 데이터를 training set(모델 구축), test set(모델 평가)로 나누어 각각의 모델에 대한 정확도 측정

(step2) 구축된 여러 모델을 통합하여 최초 통계 모델로 선택. 모형 평가란, 예측(prediction)을 위해 만든 모형이 임의 모형(random model)보다 예측력이 우수한지, 고려된 다른 모형 중 어느 모형이 가장 우수한 예측력을 보유하고 있는지를 비교, 분석하는 과정이라고 할 수 있음

(step3) 통계 모델에 의해 고위험 가구 선정 결과 list를 지자체에 제공 고위험 가구 list를 바탕으로 지자체 방문조사 수행 및 복지서비스 제공

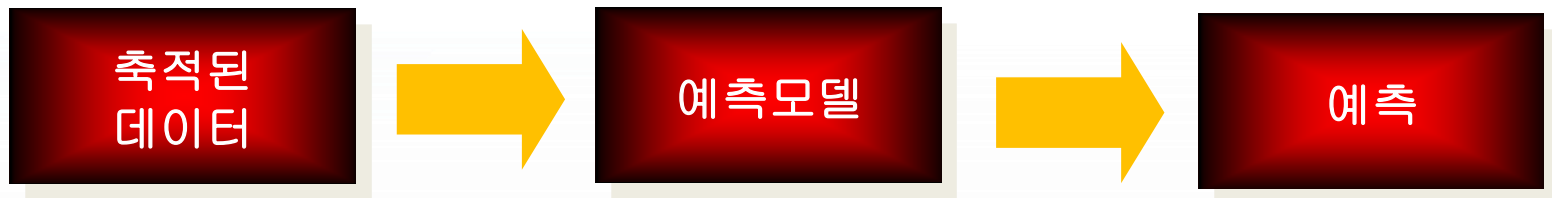
(step4) (환류) 조사 및 조치 결과를 비교 분석하여 통계 모델 개선 및 정확도 제고에 활용 ➡ 정교화된 통계 모델을 활용한 고위험 가구 선정 결과 제공 ➡ 지자체 방문조사 수행 및 결과 입력 ➡ 통계 모델 정교화 및 새로운 정보 연계 확대 (반복)

## 5. 복지사각지대 발굴 예측 모형

---

# 5 복지사각지대 위험가구 Predictive Modeling

- Predictive Modeling : 주어진 Data에 근거하여 Model을 만들고 이 Model을 이용하여 새로운 Case들에 대한 예측을 하는 작업



사각지대 대상자& 비대상자	사회보장 정보시스템 정보	실업 정보	.....	채무 정보
Y <sub>1</sub>	A <sub>1</sub>	X <sub>11</sub>	...	X <sub>1p</sub>
⋮	⋮	⋮		⋮
Y <sub>n</sub>	A <sub>n</sub>	X <sub>n1</sub>	...	X <sub>np</sub>

1. Logistic Regression
2. Logistic + Elastic Net
3. Random forest
4. Boosting

- 복지 사각지대 위험가구의 위험 정도 예측
- 복지 사각지대 발굴에 유의한 변수 정의

# 5 복지사각지대 발굴 예측 모형

## 사전조사

- ✓ 2015년 7월 시행된 사회보장급여법 및 시행령을 통해 복지 사각지대 관련 정보를 수집할 수 있는 법적 근거가 마련되었으나, 이러한 정보들이 연계되고 수집된다고 할지라도 복지 사각지대 예측을 위한 통계모형 구축을 위해서는 복지 사각지대와 관련된 특성분만 아니라 조사결과에 따른 복지 사각지대 해당 여부에 대한 정보가 필요함
- ✓ 이에 따라, ‘복지사각지대 발굴방안 마련을 위한 취약계층 실태조사’를 계획하여 수행하였음
- ✓ 이러한 사전 실태조사(일선 지자체 실태조사) 결과(6,849가구, 74개 변수)를 바탕으로 복지 사각지대 예측을 위한 분석 및 통계모형 구축을 실시
- ✓ 본 실태조사 결과를 기초로 복지사각지대 위험 확률이 높은 가구를 추출할 수 있는 초기 통계모형을 개발하고, 이를 기반으로 사각지대 발굴관리 시스템에 적용 구축함



# 5 모형 구축 프로세스

## (step1) 실태조사 기반 모형 구축 및 검증

- 1차 연구에서 수행된 복지사각지대 실태조사(6,849가구) 74개 변수 기반으로 신규 연계 정보와 동일 또는 유사한 변수 생성
- ※ 〈한계〉 실태조사 당시 복지사각지대 비율이 85.9%로 높았으며, 모형 구축 시 74개 변수 중 신규 연계정보와 동일 또는 유사한 변수만 활용
- 신규 연계 정보 변수(X)와 복지사각지대(차상위지원 대상 이하) 가능성 여부 판단 변수(Y) 기준으로, 4가지 방법론(Logistic, Elastic-net, Random forest, Boosting)별 모형 구축 및 검증

## (step2) 신규 연계 정보 활용 복지사각지대 위험가구 확률 예측

- 4개 방법론(Logistic, Elastic-net, Random forest, Boosting)별 구축 모형과 신규 연계 정보 중 해당여부 관련 변수를 활용하여 개별 사례의 잠재적 복지사각지대 위험가구 확률 예측

## (step3) 현장 발굴대상 복지사각지대 위험가구 추출 및 제공

- 개별 사례의 잠재적 복지사각지대 위험가구 확률 예측 결과를 바탕으로 현장 발굴대상 복지사각지대 위험가구 리스트 추출 및 제공
- 전체 case 중 현장 발굴대상 위험가구 규모 및 확률 순위 등을 고려하여 추출
- 복지사각지대 위험가구 확률에 따른 우선순위 및 등급 부여 후 지자체 제공

# 5 모형 구축 프로세스

## (step4) 복지사각지대 발굴 조사 및 지원결과 환류 데이터 기반 예측모형 재구축

- 실태조사의 한계를 개선하고, 연계 정보(해당 여부 및 세부정보) 활용을 통한 예측 모형 정교화를 위해 발굴 및 처리 결과(실태조사 포함) 환류 필수
- 신규 연계정보 중 다양한 상세 정보를 활용한 예측 및 모형 정교화는, 발굴대상 조사 및 지원결과 환류 데이터 기반으로 2016년 연구에서 수행
- 2015. 12월 신규 연계정보 기반 복지사각지대 위험가구 예측 결과 제공을 통해 지자체 조사 및 지원결과 환류 데이터 중심으로 모형 비교 및 재구축

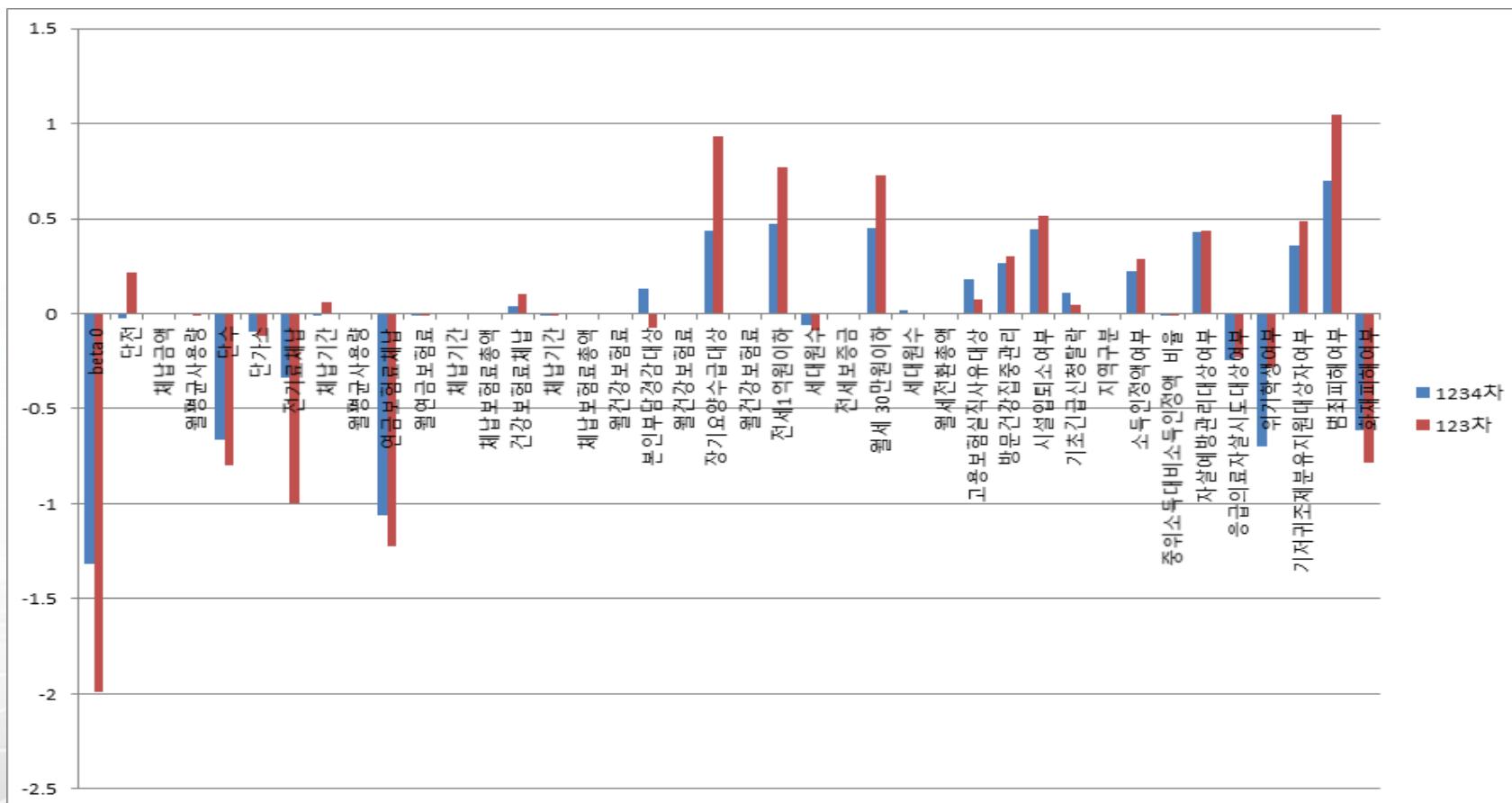
※ <비교> 실태조사 당시 조사대상의 특성으로 인해 복지사각지대 해당 비율이 85.9%로 높았으나, 1차 환류 데이터 분석대상(8,922건 중 중복지원 및 가구ID 중복을 제외한 4,785가구) 중 복지사각지대로 확인된 지원 완료가구는 20.4%로, 실태조사의 한계와 환류 데이터 분석의 중요성을 확인할 수 있음

# 5 적용 모형

method	Advantages	Disadvantages
<b>Logistic Regression</b>	<ul style="list-style-type: none"> <li>➤ 모형이 안정적</li> <li>➤ 모형의 해석이 쉬움</li> <li>➤ 계산이 어렵지 않음</li> </ul>	<ul style="list-style-type: none"> <li>➤ 오직 linear decision boundary를 제공</li> <li>➤ Interaction term 없음</li> </ul>
<b>Elastic Net</b>	<ul style="list-style-type: none"> <li>➤ 회귀분석에서 다중공선성 문제 해결 가능</li> <li>➤ 모형의 해석이 쉬움</li> <li>➤ 변수선택 가능</li> </ul>	<ul style="list-style-type: none"> <li>➤ parameter tuning</li> </ul>
<b>Random Forest</b>	<ul style="list-style-type: none"> <li>➤ 매우 높은 예측력</li> <li>➤ 독립변수의 수가 많을 때에도 좋은 결과를 얻을 수 있음</li> <li>➤ 이상치에 둔감</li> <li>➤ 독립변수의 변환에 invariant</li> <li>➤ Interaction terms : 비선형 효과 고려</li> </ul>	<ul style="list-style-type: none"> <li>➤ 이론적 설명 부족</li> <li>➤ 최종 결과에 대한 해석이 어려움</li> </ul>
<b>Boosting</b>	<ul style="list-style-type: none"> <li>➤ 매우 높은 예측력</li> <li>➤ 독립변수의 변환에 invariant</li> <li>➤ Interaction terms : 비선형 효과 고려</li> </ul>	<ul style="list-style-type: none"> <li>➤ parameter tuning</li> </ul>

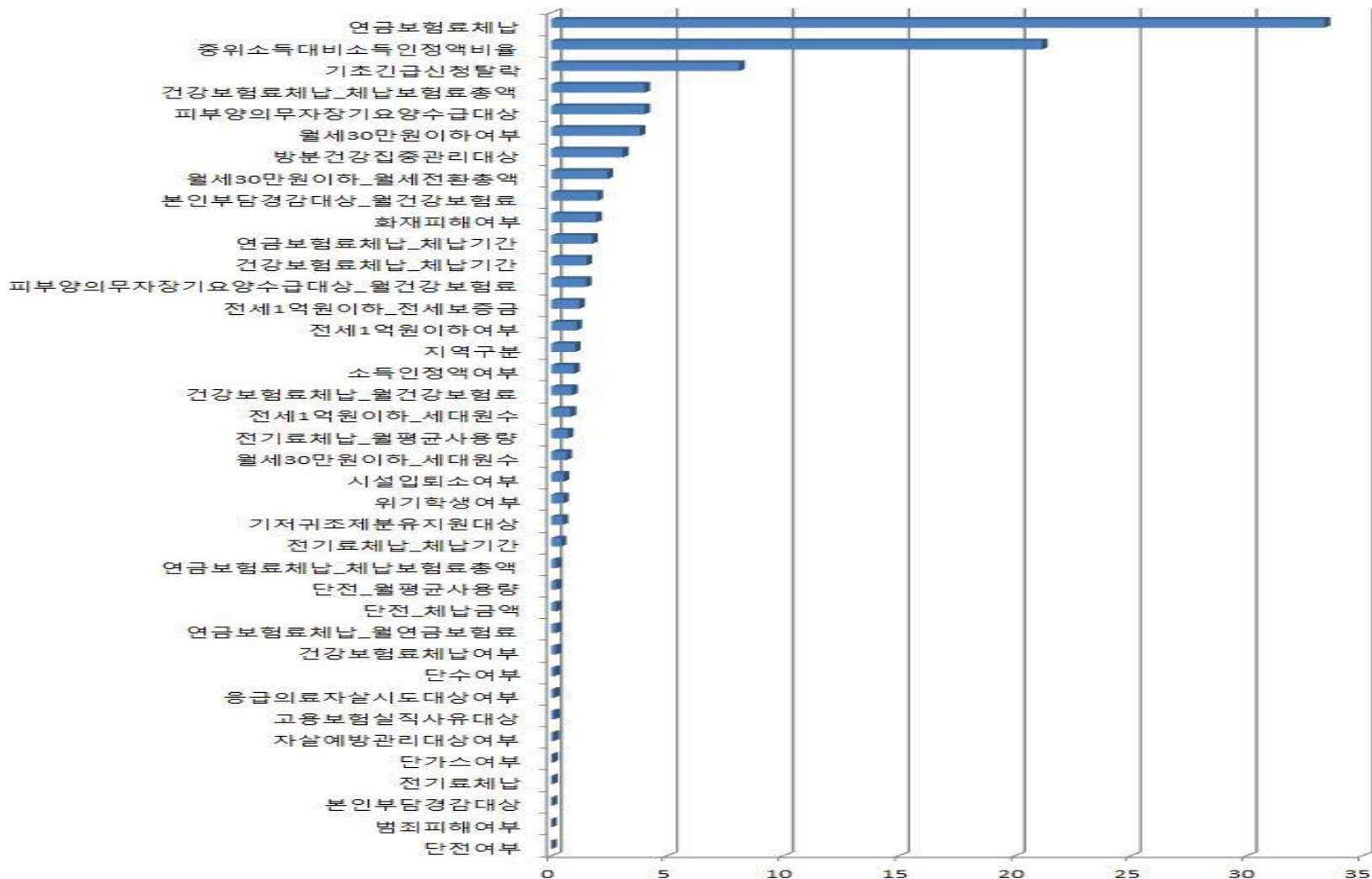
# 5 적용 모형

1~3차 환류데이터 및 1~4차 환류데이터 반영 예측모형 Logistic Coeff 변화



# 5 적용 모형

1~4차 환류데이터 반영 예측모형 Boosting 상대적 영향도



## 6. 개선 사항 및 향후 방향

---

# 6 개선 사항

- ✓ 일선 현장의 업무 담당자들이 중요하다고 생각하는 복지사각지대 발굴에 유효한 항목과 사회보장빅데이터를 활용한 예측모형에서 유의미한 항목의 간극을 줄일 필요가 있음
  - 연계정보의 유의미성에 대한 체감도 및 신규 연계정보에 대한 아이디어 수렴
- ✓ 그동안 축적된 발굴 및 지원결과에 근거하여 복지대상자 발굴 정확도 제고를 위한 통계 모형 개선
  - Deep Learning 적용 검토
- ✓ 복지사각지대 지원현황 분석 및 업무처리 개선방안 마련
  - 정보시스템을 활용하여 복지사각지대 대상자 업무처리과정이 효율적으로 이루어질 수 있도록 지원방안 마련

# 6 향후 방향

빅데이터와 인공지능 기반 복지사각지대 발굴 시스템에 의한 복지 패러다임의 전환





**감사합니다**