

빅데이터 산업에서의 정보보호 현황과 전망

장 석 호*

요 약

본고에서는 개인정보보호를 위하여 제정된 여러 관련 법률과 제도로 인하여 활성화 되지 못하는 빅데이터 산업의 현황을 금융산업을 통해 구체적으로 검토하고 있다. 그리고 현 법제 하에서 제공 가능한 데이터의 항목과 수준에 대해 제시하고자 한다. 또한 정보보호를 위한 비식별화방법과 유의해야 할 실무관점의 사항들에 대해 제언하고 있다.

I. 서 론

빅데이터는 언론에서 가장 많이 회자되는 키워드 중의 하나이다. 그럼에도 불구하고 기술의 발전이나 산업으로서 명확한 성과가 없다는 비판을 받고 있는 것도 사실이다. 그리고 그 주원인이 정보보호와 관련된다는 의견이 지배적이다.

지난 수년간 대형 정보유출 사고를 겪어 온 경험이 아직도 생생하고 이를 방지하기 위해 지속적으로 강화된 관련 법률과 제도가 산업 간 정보의 흐름을 원천적으로 차단하고 있어 다양한 형태로 가공되고 활용되지 못한다는 지적이다.

본고에서는 국내의 대표적인 빅데이터 분야인 금융영역에서의 정보보호 현황 이슈를 다루어 사안별로 검토하고 이를 발전적인 방향으로 전개할 수 있는 대안을 제시하고자 한다.

II. 금융회사에서의 빅데이터 정보보호와 관련된 법률 검토

2016년 1월 금융위원회에서는 빅데이터 산업의 활성화를 위해 관련법을 금년 안에 개정할 것을 알렸다. 각종 금융정보를 제3자 제공이 가능하도록 관련법과 유권해석을 명확히 하여 빅데이터 산업을 활성화하겠다는 취지를 알린 것이다. 이는 작년 6월 발표한 ‘빅데이터 활성화 방안’을 단계별로 이행하는 절차로 볼 수 있다. 그 주요 내용으로는 비식별정보를 신용정보에서 제외하

여 빅데이터 사업에 활용할 수 있는 근거를 마련하고 이를 핀테크 기업에도 제공할 수 있도록 한다는 것이다. 금융정보를 집중적으로 관리하는 종합신용정보 집중기관인 신용정보원을 설립하여 본격적인 빅데이터 활용 업무를 지원한다는 구체적인 방안도 제시한 바 있다.

사실, 그 이전부터 정부의 각 부처에서는 비식별화 지침을 발간하여 빅데이터를 활용할 수 있는 가이드라인을 민간에 제공하였으나 이를 근거로 빅데이터의 활용이 본격화되기에는 미흡한 면이 있어 관련 법을 개정하기로 방향을 잡는 것으로 볼 수 있다.

금융산업에서는 특별법인 신용정보법과 일반법인 개인정보보호법을 적용하며 그 주요한 내용을 아래에 검토한다. 개인정보보호법 제2조에는 개인정보를 다음과 같이 정의한다.

개인을 특정하는 구체적인 식별정보 뿐만 아니라 일부 정보라도 재식별의 가능성이 있다면 개인정보로 보는 것이며 이는 제3조 개인정보 보호 원칙에서 수집단계에서부터 처리 목적을 명확하게 밝히고 수집하여야 하며 목적 내에서만 활용하여야 한다. 개인정보는 기본

개인정보보호법 제2조 정의

“개인정보”란 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다.

* 비씨카드 빅데이터 센터장 (sokho.chang@gmail.com)

적으로 개인으로부터 동의 받은 목적 내에서만 사용될 수 있다(제15조 개인정보의 수집과 이용).

동의를 받지 않고 개인정보를 이용할 수 있는 예외에 대해서는 제 18조 개인정보의 목적 외 이용과 제공 제한에서 다루고 있다. 그 중 2항 4호에는 개인정보를 동의 받지 않고 학술적 연구의 목적으로 외부에 제공 가능한 비식별화 처리를 포함하고 있다.

개인정보보호법 제18조 2항 4호
통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

그 밖에 개인의 사상 등의 민감정보는 수집과 분석이 제한되며 주민번호와 같은 특별한 식별번호로 극도로 제한되는 것을 명확히 규정하고 있다(제23조 민감정보의 처리제한).

개인정보보호법 제23조 민감정보의 처리제한
개인정보처리자는 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보로서 대통령령으로 정하는 정보(이하 "민감정보"라 한다)를 처리하여서는 아니 된다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 그러하지 아니하다.

이제 신용정보의 이용 및 보호에 관한 법률(약칭 신용정보법)의 주요한 내용에 대해 살펴보도록 하자. 법률 제2조 정의에서의 개인신용정보는 다음과 같다.

신용정보법 제2조 정의
"개인신용정보"란 신용정보 중 개인의 신용도와 신용거래능력 등을 판단할 때 필요한 정보로서 대통령령으로 정하는 정보를 말한다.

개인정보보호법에서와 같이 개인신용정보는 수집과 제공에 대해 동의를 받는 것을 명확히 하고 있다(제32

조 개인신용정보의 제공과 활용에 대한 동의).

신용정보법 제32조 개인신용정보의 제공 활용
① 신용정보제공이용자가 대출, 보증에 관한 정보 등 대통령령으로 정하는 개인신용정보를 타인에게 제공하려는 경우에는 대통령령으로 정하는 바에 따라 해당 개인으로부터 다음 각 호의 어느 하나에 해당하는 방식으로 미리 동의를 받아야 한다.
1. 서면
2. 「전자서명법」 제2조제3호에 따른 공인전자서명이 있는 전자문서(「전자거래기본법」 제2조제1호에 따른 전자문서를 말한다)

유의할 점은 신용정보법에서는 개인정보보호법에서와 같이 동의를 받지 않고 학술연구 등의 목적 외 활용의 예외사항에 대해 규정하고 있지 않다는 것이다. 특별법인 신용정보법에 규정되어 있지 않은 사항에 대해 일반법인 개인정보보호법을 적용하는 것으로 유권해석을 한다면 가능할 수 있지만 보수적인 금융산업의 특성상 신용정보법 또한 시행령에 명확히 규정되어 근거가 마련되기를 기대하고 있다.

III. 활용 가능한 금융회사의 빅데이터

법률적 제약에도 불구하고 여전히 현행법 하에서 금융데이터를 활용할 수 있다. 활용 가능한 면이 알려져 있지 않고 법률적인 장애부분만 부각되어 제공과 활용이 전면 금지된 것으로 알려져 있어 빅데이터 산업을 더욱 퇴보시키고 있다고 생각한다.

신용카드 산업을 예를 들면, 신용카드 산업은 양면시장으로서 회원정보와 가맹점정보를 보유하고 있다. 회원정보는 개인에 관한 정보로서 개인정보보호법에 따라 활용이 지극히 제한되는 면이 있다. 그러나 가맹점 정보는 기업 또는 법인(자영업자 포함)으로 볼 것임으로 정보제공에 따른 동의 여부가 명문화된 근거법령이 없다. 개인정보가 아니라는 것이다. 그러므로 신용카드 가맹점의 거래실적에 기반한 각종 데이터는 활용 가능한 대표적 금융데이터라고 볼 수 있다. 다만, 기업의 영업비밀을 공개되지 않도록 거래금액과 건수에 대한 범주화를 하여 직접적인 영업비밀이 노출되지 않도록 하는 주

의는 필요하다.

가맹점의 매출관련 통계데이터를 고객의 특성에 관하여 세분화하여 제공할 때에도 주의가 필요하다. 예를 들어, 가맹점의 성별, 연령, 요일, 시간대별 매출을 분석할 때에 집계수량이 적어 소수집단이 포함될 경우에는 다른 정보와 결합하여 특정개인을 재식별할 가능성이 존재한다.

신용카드 회사는 방대한 회원정보도 보유하고 있다. 향후에는 법을 개정하여 카드번호, 전화번호, 주소 등의 개인식별정보를 삭제하면 동의 받지 않고 제공 등의 활용이 가능할 것이라고 밝힌 바 있다. 그러나 현재에도 개인이 식별되지 않는 그룹의 통계정보는 활용이 가능한 영역이라고 볼 수 있다는 것이 중론이다. 충분히 큰 모수의 그룹의 성별과 연령 등 인구통계학적인 특징과 행태는 개인을 특정하는 개인정보라고 볼 수 없기 때문이다.

향후에 개정될 법률의 큰 골자가 비식별화한 금융데이터를 개인정보에서 제외하여 동의 받지 않고 제공할 것인데 비식별화한 데이터는 결국 일정한 특징을 갖는 여러 가지의 그룹으로 나뉘어 범주화된 통계데이터로 가공되어 활용될 것임으로 현재의 활용 가능 내용과 크게 다르지 않을 것으로 사료된다.

혹자는 법이 개정되면 타사의 개인 고객정보를 제공받아 자사의 고객정보와 일대일 융합하여 마케팅 효율을 높이는 등의 막대한 기대를 하고 있다. 물론, 세분화된 통계정보라 하더라도 특정 개인의 행태정보보다는 가치가 낮은 것은 당연하지만 앞으로는 상당기간 빅데이터 산업활성화를 위해 개인정보보호 수준을 내릴 것으로 보는 전문가는 드물다.

그러므로 종합하여 보면, 현재 금융권에서 활용 가능한 데이터는 특정 개인의 정보가 아닌 그룹의 정보로서 총합 또는 평균 등의 총계처리(aggregation)를 하여 특정개인을 판단할 수 없도록 하고 집계수량이 적을수록 특정인을 예상하는 것이 가능할 수 있음을 유의하여야 한다. 경우에 따라서는 해당 그룹의 대표 값으로 변환하거나 구간 값으로 변환하여 고유정보 추적 및 식별 방식을 하여야 한다. 이러한 통계처리는 특정 신용정보주체의 거래내용을 판단할 수 없도록 하는 것이고 신용정보에서 제외시키는 과정이다. 법의 일관된 태도는 신용정보주체를 특정하는 경우에만 신용정보라고 보기 때문이다.

IV. 빅데이터 산업의 발전방향

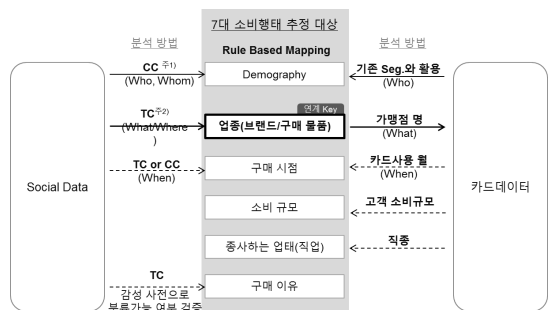
데이터 분석 솔루션은 오랜 기간 고도화 과정을 거쳐 상당한 수준에 도달했다. 또한 데이터 분석 전문가의 수도 꾸준히 증가하여 왔다. 그러나 빅데이터 산업이 더욱 활성화되기 위해서는 타 산업의 데이터가 융합되어 활용되어야 하는데 여전히 초보적인 단계에 머물러 직접적인 장애요인으로 거론된다.

세부적인 원천 데이터를 제공받아 고도의 분석과 다양한 활용도 좋지만 현 제약상황을 명확히 이해하고 법률이 허용하는 범위 내에서 적극적인 데이터의 가공을 통해 외부의 필요를 만족시키기 위해 기법과 방법론을 개발하는 면에도 역점을 기울여야 한다고 생각한다.

정밀화된 개인의 분석도 의미 있지만 외부의 데이터를 다차원으로 가공 융합하여 데이터의 가치를 높이는 노력도 지속해야 한다.

필자는 최근 각종 블로그와 커뮤니티, 뉴스 댓글 등의 소셜 데이터를 수집하여 카드 마케팅에 활용한 연구를 수행한 적이 있다. 인터넷 상의 대부분의 내용은 익명화되어 개인을 특정하기 불가능하며 문맥을 따져 추측을 할 뿐이다. 그러나 언급하는 상품의 업종을 기준으로 카드소비 데이터의 업종과 연결하여 소비행태를 여러 가지 방법론을 사용하여 분석하고 마케팅에 적용하였을 때 전통적인 마케팅의 반응성과를 상회하는 결과를 얻을 수 있어서 유용한 하나의 마케팅 기법으로 사용될 수 있음을 알 수 있었다.

소셜데이터가 부정확하고 누구인지 알 수 없어 데이터가 많지만 유용하지 않다는 의견도 있다. 그러나 행동적인 집단의 특징을 선형적으로 파악하기 위하여 소셜 데이터를 활용하는 것이 유용한 측면도 있음을 볼 수 있었다. 업종별 소비방향을 소셜데이터와 카드데이터



(그림 1) 소셜데이터와 카드데이터의 연계 활용 예시

식스포켓 산업중분류	주성분1	주성분2	주성분3	주성분4	...	주성분8
교육_도서						0.416
교육_사교육	-0.271		-0.158			-0.36
놀이체험		0.123	0.354	-0.232		0.175
유아용품		-0.106	0.322			-0.41
의류_잡화		-0.146	0.338			0.308
자전거		0.205	-0.539			
장난감				0.445		0.552
전시_공연체험		0.216	-0.406	-0.17		0.133
학용품			-0.195	-0.511		0.132
동남아			0.117	-0.644		0.158
분식	0.259	0.296	0.264			-0.124
뷔페	0.435	-0.265				
양식	0.415	-0.211				
일식	0.209	-0.353				0.107
중식	0.282	0.179				
카페_디저트	0.327	0.341				
패스트푸드	0.263	0.547				
한식	0.437	-0.267	-0.119			
해당 주성분 데이터 분산 설명 비율	19.6%	7.2%	6.2%	6%	...	5.5%
누적 비율	19.6%	26.8%	33%	39%	...	61.7%

(그림 2) 카드데이터의 주성분 분석으로 본 소비방향을

의 양쪽에서 찾아 마케팅을 위한 소비트렌드를 미리 예측하는 것이 가능하다.

이 사례처럼 외부의 여러 다양한 필요에 맞추어 다양한 형태의 통계데이터로 가공하여 준비하는 것이 산업 간 데이터 융합을 위해 반드시 준비될 사항으로 보고 있다.

V. 결 론

금융위원회가 발표한 빅데이터 활성화 방안에 따르면 이름 등 특정한 식별정보를 뺀 개인신용정보를 금융사들이 영업이나 마케팅에 자유롭게 활용할 수 있을 것이라고 하였다. 그러나 비식별화된 데이터므로 현재 활용 가능한 수준과 크게 달라지지 않을 가능성이 높다.

중요한 것은 데이터의 가공수준을 높여 직접적인 개인정보를 제공하지 않더라도 각 산업의 상품과 서비스에 유용한 다양한 고객의 특징과 양태정보를 제공할 수 있어야 한다.

빅데이터가 산업에서 본격적으로 활용되거나 수익 모델로 연결되고 있지 않아 각종 제도의 개선이 예고되고 있지만 현재 상황에서도 활용과 제공이 가능한 데이터의 수준과 범위를 명확히 하여 가공성을 높이는 노력도 병행되어야 본격적인 빅데이터로 인한 금융권과 핀테크 기업의 동반 성장도 기대할 수 있을 것이다.

참 고 문 헌

- [1] 행정자치부, “개인정보보호법”,
- [2] 금융위원회, “신용정보의 이용 및 보호에 관한 법률”
- [3] 미래창조과학부, “빅데이터 활용을 위한 개인정보 비식별화 사례집”, 2014.5

<저자소개>



장 석 호 (Chang Sokho)

2005년 8월 : 서울대학교 경영학과 석사

1998년 2월 : 강원대학교 전자공학과 박사

1998년 3월~2000년 2월 : 한국과학기술연구원 Post Doc

현재 : 비씨카드 빅데이터 센터장

<관심분야> 빅데이터, 핀테크, 금융정보보호