

## 【발명의 설명】

### 【발명의 명칭】

공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템{SYSTEM FOR PROVIDING BIGDATA BASED RESERVATION PRICE PROBABILITY DISTRIBUTION PREDICTION SERVICE FOR PROCUREMENT AUCTION}

### 【기술분야】

<0001> 본 발명은 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템에 관한 것으로, 빅데이터를 분석하여 가능한 예정가격의 확률분포를 시각화하여 제공할 수 있는 플랫폼을 제공한다.

### 【발명의 배경이 되는 기술】

<0002> 한국의 공공조달시장 규모는 전체 세출 예산의 1/3 수준에 이르고 있어 국민 경제에서 차지하는 비중과 영향력이 매우 크다. 특히, 공공조달시장은 대금 지급이 빠르고 현금으로 결제되는 특성이 있어 조달시장 참여 업체 입장에서는 안정적 판로 확보 측면에서 낙찰자로 선정되어 공공조달을 수주하는 것이 매우 중요하며, 정부 측 입장에서는 정해진 예산의 범위 내에서 필요한 물품이나 공사 및 용역 등을 조달해야 하므로 적정한 가격으로 입찰한 자를 낙찰자로 선정하는 것이 매우 중요한 문제가 된다. 정부나 공공기관이 물품을 조달하거나 공공사업을 시행할 때 경쟁입찰을 통해 가장 낮은 가격의 입찰자를 낙찰자로 선정하나, 가격 이외에 업체의 능력이나 재화 및 서비스의 품질 등을 고려하여 적격한 업체를 낙찰자로 정하는 등 다양한 형태의 조달 경매(Procurement Auction) 방식이 사용되고 있다. 한국의

경우 경쟁 입찰을 통한 낙찰자 선정 방식을 관련 법령에 규정하고 있는데, 이에 따르면 공공조달 시 국고의 부담이 되는 경쟁 입찰의 경우 예정가격 이하로서 최저가격으로 입찰한 순서대로 계약이행 능력 등을 심사하여 낙찰자를 결정하는 적격심사 제도를 가장 우선적으로 적용하도록 하고 있다.

<0003> 이때, 공공조달시장에서 전자입찰방식 및 빅데이터를 분석하여 자동으로 입찰하는 방법이 연구 및 개발되었는데, 이와 관련하여 선행기술인 한국공개특허 제2003-0030140호(2003년04월18일 공개) 및 한국공개특허 제2017-0069907호(2017년06월21일 공개)에는, 정부, 공공기관 및 기업 간 전자상거래에서 인터넷을 이용한 전자입찰 및 서류입찰을 동시에 처리하고, 입찰관련 수수료 및 입찰보증금을 공인인증기관의 인증에 의하여 자동이체되도록 하고, 조달청에서 실시하고 있는 입찰정보, 규격서 작성, 공동수급협정서, 입찰서 및 공고서를 포함하는 입찰 프로세스가 인증 및 보안을 통하여 실행가능하도록 하는 구성과, 입찰서버로부터 데이터를 수집하고, 사용자 단말에서 옵션정보를 선택하면 빅데이터 기반 입찰시점 및 예측입찰가를 연산하여 견적입찰가를 제공하는 구성이 각각 개시되어 있다.

<0004> 다만, 전자의 경우 오프라인에서 진행되던 입찰을 전자입찰로 변경한 것 이상의 의미가 없고, 후자의 경우에도 광고 노출순위나 광고 입찰가를 예측한 구성으로 공공조달의 입찰과는 거리가 있다. 한국의 적격심사제도는 지나친 가격경쟁으로 인해 발생할 수 있는 납품 물품이나 공사 및 용역의 품질 하락 등을 예방하기 위해 일정 점수 이상을 획득한 업체를 낙찰자로 선정하도록 하고 있으며, 입찰 참여 업체 입장에서 자신의 비가격 요소 점수가 다른 업체와 동일하다고 판단될 경우

낙찰자가 되기 위해서는 일정 점수 이상을 얻을 수 있는 수준으로 입찰가격을 제시하여야 한다. 이에, 적격심사제도에서 물품생산 능력 등의 비가격 요소 점수가 일정 조건을 만족할 경우 입찰참여 업체가 제시한 입찰가격이 낙찰자 결정에 매우 중요한 역할을 하게 된다. 따라서, 빅데이터 기반으로 예정가격의 확률분포를 제공함으로써 의사결정권자가 최종 예정가격을 결정할 수 있도록 의사지원을 제공하는 플랫폼의 연구 및 개발이 요구된다.

**【발명의 내용】**

**【해결하고자 하는 과제】**

<0005>

본 발명의 일 실시예는, 웹크롤링을 통하여 빅데이터를 수집하고, 수집된 빅데이터 내 포함된 반정형 데이터 및 비정형 데이터를 실시간 스트림(Realtime Stream)처리하여 RDB(Relational Database)에 적재하며, RDB에 적재된 데이터가 정부입찰 및 계약 집행기준과 어긋난 오류를 자동으로 찾아 수정하는 정제 및 표준화 작업을 실행함으로써 전처리하고, 공공입찰의 예정가격 확률분포 정보를 의사결정권자에게 제공함으로써 비즈니스 성과를 달성할 수 있도록 도와줄 수 있는, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법을 제공할 수 있다. 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

**【과제의 해결 수단】**

<0006>

상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 실시예는, 공고목록에서 입찰가격을 산출하고자 하는 공고번호를 선택하고, 선택된 공

고번호의 공공입찰에 참여하는 경우 낙찰가능한 적어도 하나의 예정가격(Reservation Price)을 빅데이터 기반 데이터 분석으로 추출하여 출력하는 사용자 단말 및 적어도 하나의 공공조달입찰 데이터를 웹크롤러를 이용하여 수집하는 수집부, 수집된 적어도 하나의 공공조달입찰 데이터 내 정형(Structured) 데이터, 반정형(Semi-Structured) 및 비정형(Unstructured) 데이터를 적재하는 적재부, 적어도 하나의 공공조달입찰 데이터가 정부입찰 집행기준 또는 정부계약 집행기준과 어긋난 오류(Error)를 찾아 수정, 정정 및 삭제 중 어느 하나 또는 적어도 하나의 조합에 대응하는 정제 및 표준화 절차를 포함하는 전처리를 수행하는 전처리부, 사용자 단말에서 선택한 공공입찰의 적어도 하나의 예정가격을 확률분포로 시각화하여 사용자 단말로 전달하는 분석부를 포함하는 예측 서비스 제공 서버를 포함한다.

#### 【발명의 효과】

<0007>        전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 웹크롤링을 통하여 빅데이터를 수집하고, 수집된 빅데이터 내 포함된 반정형 데이터 및 비정형 데이터를 실시간 스트림(Realtime Stream)처리하여 RDB(Relational Database)에 적재하며, RDB에 적재된 데이터가 정부입찰 및 계약 집행기준과 어긋난 오류를 자동으로 찾아 수정하는 정제 및 표준화 작업을 실행함으로써 전처리하고, 공공입찰의 예정가격 확률분포 정보를 의사결정권자에게 제공함으로써 비즈니스 성과를 달성할 수 있도록 도와줄 수 있다.

#### 【도면의 간단한 설명】

<0008>        도 1은 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터

기반 예정가격 확률분포 예측 서비스 제공 시스템을 설명하기 위한 도면이다.

도 2는 도 1의 시스템에 포함된 예측 서비스 제공 서버를 설명하기 위한 블록 구성도이다.

도 3 및 도 4는 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스가 구현된 일 실시예를 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법을 설명하기 위한 동작 흐름도이다.

**【발명을 실시하기 위한 구체적인 내용】**

<0009>

아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

<0010>

명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그

이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

<0011> 명세서 전체에서 사용되는 정도의 용어 "약", "실질적으로" 등은 언급된 의미에 고유한 제조 및 물질 허용오차가 제시될 때 그 수치에서 또는 그 수치에 근접한 의미로 사용되고, 본 발명의 이해를 돕기 위해 정확하거나 절대적인 수치가 언급된 개시 내용을 비양심적인 침해자가 부당하게 이용하는 것을 방지하기 위해 사용된다. 본 발명의 명세서 전체에서 사용되는 정도의 용어 "~(하는) 단계" 또는 "~의 단계"는 "~를 위한 단계"를 의미하지 않는다.

<0012> 본 명세서에 있어서 '부(部)'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1 개의 유닛이 2 개 이상의 하드웨어를 이용하여 실현되어도 되고, 2 개 이상의 유닛이 1 개의 하드웨어에 의해 실현되어도 된다. 한편, '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, '~부'는 어드레싱 할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체 지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 '~부'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성

요소들과 '~부'들로 더 분리될 수 있다. 뿐만 아니라, 구성요소들 및 '~부'들은 디바이스 또는 보안 멀티미디어카드 내의 하나 또는 그 이상의 CPU들을 재생시키도록 구현될 수도 있다.

<0013> 본 명세서에 있어서 단말, 장치 또는 디바이스가 수행하는 것으로 기술된 동작이나 기능 중 일부는 해당 단말, 장치 또는 디바이스와 연결된 서버에서 대신 수행될 수도 있다. 이와 마찬가지로, 서버가 수행하는 것으로 기술된 동작이나 기능 중 일부도 해당 서버와 연결된 단말, 장치 또는 디바이스에서 수행될 수도 있다.

<0014> 본 명세서에서 있어서, 단말과 매핑(Mapping) 또는 매칭(Matching)으로 기술된 동작이나 기능 중 일부는, 단말의 식별 정보(Identifying Data)인 단말기의 고유번호나 개인의 식별정보를 매핑 또는 매칭한다는 의미로 해석될 수 있다.

<0015> 이하 첨부된 도면을 참고하여 본 발명을 상세히 설명하기로 한다.

<0016> 도 1은 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템을 설명하기 위한 도면이다. 도 1을 참조하면, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템(1)은, 적어도 하나의 사용자 단말(100), 예측 서비스 제공 서버(300), 적어도 하나의 관리자 단말(400)을 포함할 수 있다. 다만, 이러한 도 1의 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템(1)은, 본 발명의 일 실시예에 불과하므로, 도 1을 통하여 본 발명이 한정 해석되는 것은 아니다.

<0017> 이때, 도 1의 각 구성요소들은 일반적으로 네트워크(Network, 200)를 통해

연결된다. 예를 들어, 도 1에 도시된 바와 같이, 적어도 하나의 사용자 단말(100)은 네트워크(200)를 통하여 예측 서비스 제공 서버(300)와 연결될 수 있다. 그리고, 예측 서비스 제공 서버(300)는, 네트워크(200)를 통하여 적어도 하나의 사용자 단말(100), 적어도 하나의 관리자 단말(400)과 연결될 수 있다. 또한, 적어도 하나의 관리자 단말(400)은, 네트워크(200)를 통하여 예측 서비스 제공 서버(300)와 연결될 수 있다.

<0018> 여기서, 네트워크는, 복수의 단말 및 서버들과 같은 각각의 노드 상호 간에 정보 교환이 가능한 연결 구조를 의미하는 것으로, 이러한 네트워크의 일 예에는 근거리 통신망(LAN: Local Area Network), 광역 통신망(WAN: Wide Area Network), 인터넷(WWW: World Wide Web), 유무선 데이터 통신망, 전화망, 유무선 텔레비전 통신망 등을 포함한다. 무선 데이터 통신망의 일례에는 3G, 4G, 5G, 3GPP(3rd Generation Partnership Project), 5GPP(5th Generation Partnership Project), LTE(Long Term Evolution), WIMAX(World Interoperability for Microwave Access), 와이파이(Wi-Fi), 인터넷(Internet), LAN(Local Area Network), Wireless LAN(Wireless Local Area Network), WAN(Wide Area Network), PAN(Personal Area Network), RF(Radio Frequency), 블루투스(Bluetooth) 네트워크, NFC(Near-Field Communication) 네트워크, 위성 방송 네트워크, 아날로그 방송 네트워크, DMB(Digital Multimedia Broadcasting) 네트워크 등이 포함되나 이에 한정되지는 않는다.

<0019> 하기에서, 적어도 하나의 라는 용어는 단수 및 복수를 포함하는 용어로 정의

되고, 적어도 하나의 라는 용어가 존재하지 않더라도 각 구성요소가 단수 또는 복수로 존재할 수 있고, 단수 또는 복수를 의미할 수 있음은 자명하다 할 것이다. 또한, 각 구성요소가 단수 또는 복수로 구비되는 것은, 실시예에 따라 변경가능하다 할 것이다.

<0020> 적어도 하나의 사용자 단말(100)은, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 관련 웹 페이지, 앱 페이지, 프로그램 또는 애플리케이션을 이용하여 공공입찰에 참여하고자 하는 개인 및 기업의 단말일 수 있다. 이때, 사용자 단말(100)은 입찰에 참여하고자 하는 공고번호를 선택한 후, 예측 서비스 제공 서버(300)에서 예측한 예정가격을 수신하는 단말일 수 있다.

<0021> 여기서, 적어도 하나의 사용자 단말(100)은, 네트워크를 통하여 원격지의 서버나 단말에 접속할 수 있는 컴퓨터로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 네비게이션, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(Desktop), 랩톱(Laptop) 등을 포함할 수 있다. 이때, 적어도 하나의 사용자 단말(100)은, 네트워크를 통해 원격지의 서버나 단말에 접속할 수 있는 단말로 구현될 수 있다. 적어도 하나의 사용자 단말(100)은, 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, 네비게이션, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless

Broadband Internet) 단말, 스마트폰(Smartphone), 스마트 패드(Smartpad), 태블릿 PC(Tablet PC) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.

<0022>           예측 서비스 제공 서버(300)는, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 웹 페이지, 앱 페이지, 프로그램 또는 애플리케이션을 제공하는 서버일 수 있다. 그리고, 예측 서비스 제공 서버(300)는, 적어도 하나의 기관에서 공시한 데이터를 웹크롤러 또는 봇(Bot)을 이용하여 수집하는 서버일 수 있다. 또한, 예측 서비스 제공 서버(300)는 빅데이터를 구축하기 위하여 데이터를 분류하고, 전처리를 수행함으로써 오류를 제거하고 표준화를 해준 후, 빅데이터 분석, 데이터마이닝 및 역확률 모델 중 어느 하나 또는 적어도 하나의 조합을 이용하여 예정가격을 추출하여 사용자 단말(100)로 전송하는 서버일 수 있다.

<0023>           여기서, 예측 서비스 제공 서버(300)는, 네트워크를 통하여 원격지의 서버나 단말에 접속할 수 있는 컴퓨터로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 네비게이션, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(Desktop), 랩톱(Laptop) 등을 포함할 수 있다.

<0024>           적어도 하나의 관리자 단말(400)은, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 관련 웹 페이지, 앱 페이지, 프로그램 또는 애플리케이션을 이용하여 빅데이터를 구축할 때 데이터에 오류가 있지 않은지를 확인하고 수정, 삭제, 정정하고, 데이터의 형태가 표준화되도록 하는 관리자의 단말일 수 있다.

<0025>

여기서, 적어도 하나의 관리자 단말(400)은, 네트워크를 통하여 원격지의 서버나 단말에 접속할 수 있는 컴퓨터로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 네비게이션, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(Desktop), 랩톱(Laptop) 등을 포함할 수 있다. 이때, 적어도 하나의 관리자 단말(400)은, 네트워크를 통해 원격지의 서버나 단말에 접속할 수 있는 단말로 구현될 수 있다. 적어도 하나의 관리자 단말(400)은, 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, 네비게이션, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet) 단말, 스마트폰(Smartphone), 스마트 패드(Smartpad), 타블렛 PC(Tablet PC) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.

<0026>

도 2는 도 1의 시스템에 포함된 예측 서비스 제공 서버를 설명하기 위한 블록 구성도이고, 도 3 및 도 4는 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스가 구현된 일 실시예를 설명하기 위한 도면이다.

<0027>

도 2를 참조하면, 예측 서비스 제공 서버(300)는, 수집부(310), 적재부(320), 전처리부(330), 분석부(340), 빅데이터화부(350) 및 역확률모델부(360)를

포함할 수 있다.

<0028> 본 발명의 일 실시예에 따른 예측 서비스 제공 서버(300)나 연동되어 동작하는 다른 서버(미도시)가 적어도 하나의 사용자 단말(100) 및 적어도 하나의 관리자 단말(400)로 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 애플리케이션, 프로그램, 앱 페이지, 웹 페이지 등을 전송하는 경우, 적어도 하나의 사용자 단말(100) 및 적어도 하나의 관리자 단말(400)은, 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 애플리케이션, 프로그램, 앱 페이지, 웹 페이지 등을 설치하거나 열 수 있다. 또한, 웹 브라우저에서 실행되는 스크립트를 이용하여 서비스 프로그램이 적어도 하나의 사용자 단말(100) 및 적어도 하나의 관리자 단말(400)에서 구동될 수도 있다. 여기서, 웹 브라우저는 웹(WWW: World Wide Web) 서비스를 이용할 수 있게 하는 프로그램으로 HTML(Hyper Text Mark-up Language)로 서술된 하이퍼 텍스트를 받아서 보여주는 프로그램을 의미하며, 예를 들어 넷스케이프(Netscape), 익스플로러(Explorer), 크롬(Chrome) 등을 포함한다. 또한, 애플리케이션은 단말 상의 응용 프로그램(Application)을 의미하며, 예를 들어, 모바일 단말(스마트폰)에서 실행되는 앱(App)을 포함한다.

<0029> 도 2를 설명하기 이전에 본 발명의 일 실시예의 서비스를 이해하기 위한 기본개념을 먼저 설명하기로 한다. 이하에 설명된 기본개념은 도 2에서 중복하여 설명되지 않는다.

<0030> 한국은 경쟁 입찰을 통한 낙찰자 선정 방식을 관련 법령인 국가를 당사자로

하는 계약에 관한 법률 시행령 제42조 제1항에 규정하고 있는데, 이에 따르면 공공 조달 시 국고의 부담이 되는 경쟁 입찰의 경우 예정가격, 즉 공공조달의 계약담당자가 입찰 또는 계약체결 전에 낙찰 및 계약금액의 결정된 예정가격 이하로서 최저 가격으로 입찰한 순서대로 계약이행 능력 등을 심사하여 낙찰자를 결정하는 적격심사제도를 가장 우선적으로 적용하도록 하고 있다. 보다 구체적으로는 물품생산(공사) 능력 등의 비가격 요소를 평가한 점수와 입찰가격을 평가한 가격점수를 종합적으로 심사하여 낙찰자로 결정하는 방식이다. 다만, 적격심사제도는 지나친 가격경쟁으로 인해 발생할 수 있는 납품 물품이나 공사 및 용역의 품질 하락 등을 예방하기 위해 일정 점수 이상을 획득한 업체를 낙찰자로 선정하도록 하고 있다. 이에 따라 입찰 참여 업체 입장에서 자신의 비가격 요소 점수가 다른 업체와 동일하다고 판단될 경우 낙찰자가 되기 위해서는 일정 점수 이상을 얻을 수 있는 수준, 즉 예정가격 대비 낙찰 받을 수 있는 최저가격의 비율인 낙찰하한율로 입찰가격을 제시하여야 한다. 따라서 한국의 적격심사제도에서 물품생산(공사) 능력 등의 비가격 요소 점수가 일정 조건을 만족할 경우 입찰참여 업체가 제시한 입찰가격이 낙찰자 결정에 매우 중요한 역할을 하게 된다.

<0031>

현재 한국의 조달시장에서 예정가격은 개찰 전까지 비공개하는 것을 원칙으로 하고 있으며, 누설 등에 따른 비리 문제를 예방하기 위해 기초금액, 즉 예정가격을 결정하기 위해 가격조사 또는 원가 계산 방식으로 산정한 금액으로, 공사의 경우 발주처에서 조사한 당해 공사의 공사금액의 일정 범위 내에서 15개의 예비가격을 무작위로 선정하여 그중에 추첨된 4개의 예비가격을 산술평균하여 예정가격을

결정하고 있다. 이 경우 어떤 예정가격이 선정될지는 입찰자뿐 아니라 발주자(기관)도 개찰 전까지 알 수 없다. 따라서 예정가격 이하로 최저가격으로 입찰한 순서에 따라 낙찰 여부를 심사하는 한국의 적격심사제도의 특성상 예정가격의 공개 여부는 입찰참가자가 어떤 가격으로 입찰에 참여할지를 결정하는 데 있어 매우 중요하게 작용한다.

<0032>

이와 관련하여 예정가격의 공개 여부가 입찰참가자가 제시하는 입찰가격 및 입찰 결과에 미치는 영향과 관련한 연구가 국내외적으로 진행된 바 있는데, 이러한 연구들은 연구가설 등에 따라 다소 상반된 분석 결과를 제시하고 있다. Elyakime, Laffont, Loisel, and Vuong의 경우, 프랑스 정부가 입목(Standing Timber) 경매에서 이용하는 최고가격 밀봉 입찰(First-Price Sealed Bid Auction)에 대해 분석한 바 있는데, 그들은 경쟁 입찰에서 판매자가 단일한 예정가격(Reservation Price)을 입찰자에게 비공개하는 경우에 대해 모형을 설정하고 베이지안 균형 전략을 구하였다. 그들은 여기서 예정가격을 공개하는 전략이 비공개하는 전략보다 판매자에게 좋다는 것을 이론적 및 실증적으로 보인 바 있다.

<0033>

Tan은 독립적 확률분포와 사적 가치(Private Values)를 갖는 모형에서 위험 중립적인 구매자(정부)가 많은 공급업체들과 조달계약을 체결하는 것을 분석하였다. 그는 Elyakime, Laffont, Loisel, and Vuong와 같이 구매자가 계약을 제안할 때 미리 예정가격을 공급업체들에 공고하여 자신의 수요 정보를 나타내고 최고가격 밀봉 입찰을 사용하는 것이 최적 조달 메커니즘임을 보였다. 이때, 사적 가치란 입찰참가자가 얻는 가치가 입찰에 참가하는 개인마다 다르며, 각 참가자는 자기 자

신에 대한 가치는 알지만 다른 참가자에 대한 가치는 정확히 모르는 경우를 말한다. 따라서 사적 가치의 가정에서 입찰자의 수익은 자신이 알고 있는 정보, 낙찰 받을지 여부와 지불할 금액에만 의존하게 된다. 반면에 입찰참가자가 얻는 가치가 모두 동일하며 각참가자는 자신에 대한 가치뿐만 아니라 다른 참가자에 대한 가치도 알고 있는 경우인 공동가치의 가정에서는 입찰자의 수익이 자신뿐만 아니라 다른 입찰자들의 정보와 취향과 경매에 참여하지 않은 사람들의 선호에도 의존한다.

<0034>

한편, 공급자가 위험 기피적(Risk Averse)일 경우에는 공개된 예정가격을 갖는 최저가격 밀봉 입찰은 최적 메커니즘이 아닐 수 있다는 점도 보였는데, 이는 예정가격을 감추게 되면 그렇지 않은 경우에 비해 위험 기피적인 공급자들이 더 공격적으로 입찰하도록 유도되기 때문인 것으로 분석하였다. 반면, Vicent의 경우에는 입찰자들이 공동 가치(Common Value)를 갖는 경매에서 확률적 유보 가치(Reservation Value)를 갖는 판매자가 예정가격을 비공개할 때 공개할 때보다 더 높은 사전적 기대효용을 얻을 수 있다는 것을 보였다. 이는 예정가격을 비공개함으로써 공개했을 때는 참여하지 않았을 예정가격 미만의 구매 희망자들이 입찰에 참여하기 때문인 것으로 분석된다.

<0035>

Rosar의 경우 불확실성이 존재하는 복수의 예정가격을 입찰에 사용하는 경우를 분석하였는데, 그는 경매자가 판매자로서 물건을 판매할 때 경매가 실제 발생하기 전에 경매 규칙을 정하고 공고하는 기간에 주목했다. 경매 규칙을 공고하는 기간 동안 잠재적 구매자는 입찰을 준비하고 판매자는 경매물에 대한 자신의 이용 가치 등에 대해 더 많은 정보를 얻을 수 있으므로, 사전적 관점에서 판매자는 자신의

경매물의 가치가 낮을 때보다 높을 때 판매하는 것이 더 높은 효용을 얻을 수 있다. 이에 따라 판매자는 자신의 경매물에 대한 정보를 얻기 전에 사전에 공개된 예정가격을 사용하는 것보다 이러한 정보를 얻은 후 예정가격을 설정할 수 있다면 그렇게 하려고 할 유인이 발생하게 된다. Rosar는 이러한 경매자의 유인에 주목하여 최고가격 밀봉입찰에서 경매 규칙을 사전에 공지하고 추후에 예정가격을 설정할 권리를 갖는 것이 위험 기피적 판매자에게 최적임을 보였다. 또한 구매자에게 중간 수준 가격대에 입찰을 금지시키고 극단적 입찰을 선택하도록 할 때도 최적임을 보였다.

<0036>

한편, 한국의 적격심사제도의 낙찰자 결정과 관련한 이론적 연구로 김봉주의 연구가 있는데, 베이저안 균형(Bayesian Equilibrium)의 개념을 이용하여 입찰자들의 가격 전략과 그 균형의 특성을 분석하였다. 그는 독립된 확률분포를 갖는 사적 가치 경매 모형에서 구매자의 기대비용 관점에서 Elyakime, Laffont, Loisel, and Vuong와 마찬가지로 예정가격을 공개하는 것이 예정가격을 비공개할 때보다 정책의 성과가 좋을 수 있음을 보였다. 다만, 김봉주의 연구는 입찰참가자의 기대이득을 계산함에 있어 매개변수의 특정 값에 의존하고 있어 일반적 결론을 도출하는 데 있어 한계가 있으며, 낙찰자 선정을 위해 불확실성이 존재하는 복수의 예정가격을 활용하는 경우에 대한 분석도 엄밀하지 못한 한계가 있다. 이 외에도, 한국등록특허 제10-1005050호(2010년12월30일 공고), 조달청 나라장터에 복수예비가격 및 예정가격의 산출원리 등에 기본개념이 상세히 개시되어 있으므로 이를 참조하도록 한다.

<0037>

이에, 본 발명의 일 실시예는, 웹크롤러를 이용하여 국가기관, 자치단체 및

공공기관에서 제공하는 조달 데이터를 수집 및 변환하고, 수집 및 변환된 반정형 또는 비정형 데이터를 실시간 스트림(Real-time Stream)으로 처리하여 RDB(Relational Database)에 적재하고, RDB 내 정량적 데이터 및 정성적 데이터가 집행기준과 어긋난 오류(Error)를 자동으로 수정함으로써 정제 및 표준화를 거치는 전처리를 실행하고, 빅데이터를 기반으로 랜덤추첨에 따른 예약가격(Reservation Price) 확률분포를 예측하는 모델을 모델링함으로써, 의사결정권자에게 빠르고 정확한 의사결정을 지원 및 유도할 수 있는 방법을 제공하기로 한다.

<0038> 도 2를 참조하면, 수집부(310)는, 적어도 하나의 공공조달입찰 데이터를 웹 크롤러를 이용하여 수집할 수 있다. 수집부(310)는, 매크로 봇(Macro Bot) 또는 웹크롤러를 이용할 수 있는데, 매크로 프로그램은 단순/반복적 작업을 자동으로 프로그래밍하여 처리하는 소프트웨어로 통상 수작업을 통해 이루어지는 키보드 및 마우스 입력값, 입력순서 등의 작업을 사전에 입력하여 자동적, 반복적으로 실행하도록 함으로써 컴퓨터를 이용한 단순작업의 효율성을 증대시키는 데 사용된다. 매크로 프로그램을 이용하여 단순/반복적 작업을 자동화하는 것 자체는 효율적인 업무 처리를 위하여 필요하고, 아울러 매크로 프로그램은 허용된 명령을 자동화하는 방식으로 구성될 수 있다.

<0039> 또는 수집부(310)는 웹크롤러를 이용할 수도 있는데, 웹크롤러란 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하는 컴퓨터 프로그램이다. 웹크롤러가 하는 작업을 웹크롤링(WebCrawling) 혹은 스파이더링(Spidering)이라 부르며 봇이나 소프트웨어 에이전트의 한 형태이다. 웹크롤러는 크게 일반 웹크롤러와 분산 웹크

롤러가 있다. 웹크롤러의 기본 동작을 설명하면, 우선 URL 프론티어(Frontier) 모듈에서 URL을 가져와 HTTP 프로토콜을 사용해 해당 URL의 웹 페이지를 가져오는 것으로 시작한다. 그런 다음 패치(Fetch) 모듈에서 임시 저장소에 웹 페이지를 저장하고, 파서(Parser) 모듈에서 텍스트와 링크를 추출을 하고 텍스트는 인덱서(Indexer)에 보내진다. 링크의 경우는 URL 프론티어에 추가되어야 하는지에 대해 Content Seen, URL Filter, Duplication URL Element 모듈들을 거치면서 판단하게 된다. 이때, 웹 문서를 전부를 일반 웹크롤러로 크롤링 한다는 것은 사실상 불가능하기 때문에 분산 웹크롤러를 더 사용할 수 있다.

<0040>

분산 웹크롤러는 크게 2가지로 나누어지는데 그 중 하나가 중앙 집중식(Centralized) 방식이고 다른 하나는 P2P(or Fully-Distributed) 방식이다. 중앙 집중식 분산 웹크롤러는 URL 매니저가 서버와 같은 역할을 수행하고, 크롤러가 클라이언트 역할을 하는 구조이다. 크롤러에서 문서를 다운로드 받고 아웃링크(OutLink) URL을 추출하여 URL 매니저에게 넘겨주면 URL 매니저는 다운로드 받은 문서의 URL인지 검사하여 URL 중복을 제거를 한다. 즉 일반 웹크롤러에서 URL 중복과 URL 관리를 하는 부분을 URL 매니저가 대신 해 주는 것이다. 한편, P2P 방식은 각 Crawler가 완전 독립적인 구조를 가진다. P2P 방식은 각각의 크롤러가 일반 웹크롤러처럼 동작을 한다. 각각의 크롤러는 문서를 다운로드 받고 OutLink URL을 추출하고 URL 중복제거까지 모두 각각의 크롤러가 독립적으로 동작한다. 이렇게 하기 위해서는 각각의 크롤러에서 관리하는 다운로드 받은 URL 목록은 서로 배타적이어야 한다. 그렇지 않으면 서로 다른 크롤러에서 같은 문서를 다운로드 받는 현

상이 발생할 것이다. 이것을 해결하는 방법으로 각각의 크롤러는 다운로드 받을 URL 도메인(Domain)을 서로 배타적으로 나눠서 관리할 수 있다. 즉, 자신이 다운로드 도메인에 속하는 것만 관리하고 나머지 URL은 다른 크롤러에게 넘기는 방법인데, 이러한 방법을 이용하는 경우 각각의 크롤러가 독립적으로 동작할 수 있다.

<0041>       그 다음은 웹 콘텐츠를 추출해야 하는데, 웹 콘텐츠 추출 기술은 웹 문서로부터 정보 분석에 활용될 콘텐츠 등을 자동으로 추출하는 기능을 제공한다. 웹 콘텐츠 추출 시스템은 콘텐츠를 추출하는 규칙을 자동 생산해 콘텐츠만을 추출하는 장치로 콘텐츠 추출 규칙을 자동 생성하는 규칙 생성기(Rule Generator), 주어진 웹 문서에서 내비게이션 콘텐츠를 제거하는 내비게이션 콘텐츠 제거기(Navigation Content Eliminator), 콘텐츠 추출 규칙 키워드 유사도 비교를 통해 콘텐츠를 추출하는 콘텐츠 추출기(Core Context Extractor)로 구성될 수 있다. 물론, 상술한 크롤링 방법 이외에도 다양한 방법으로 실시간 자료를 수집하는 것도 가능하며 나열된 것들로 한정되지 않으며 열거되지 않은 이유로 배제되지 않는다.

<0042>       적재부(320)는, 수집된 적어도 하나의 공공조달입찰 데이터 내 정형(Structured) 데이터, 반정형(Semi-Structured) 및 비정형(Unstructured) 데이터를 적재할 수 있다. 적재부(320)는, 정형 데이터는 RDBMS(Relational DataBase Management System) 데이터베이스 저장방식으로 MSSQL(Microsoft SQL)에 저장하고, 반정형 데이터는 RDB(Relational Database)에 저장하고, 비정형 데이터는, 분산파일시스템인 DFS(Distributed File System)에 저장할 수 있다. 적재부(320)는, 적어도 하나의 공공조달입찰 데이터를 실시간 스트림(Realtime Stream) 처리하여 적

재할 수 있다. 이때, 정형 데이터는 데이터규격과 형태가 정해져 있어 고정된 필드에 저장되는 데이터이고, 반정형 데이터는, 고정된 형태는 아니지만, 일종의 스키마를 포함하는 데이터이고, 비정형 데이터는, 형태가 불규칙한 데이터이다. 이렇게 데이터의 종류나 규격이 서로 다르기 때문에 각각 서로 다른 데이터베이스에 저장을 하게 되고, 이렇게 저장만 된 데이터를 로우 데이터(Raw Data)라고 하며, 이후 오류 제거 및 표준화 작업인 전처리 작업을 거쳐 예정가격을 예측할 수 있는 근거인 빅데이터로 구축되게 된다.

<0043> 적재부(320)는, 예를 들어, 분석을 위한 데이터의 수집을 위해 조달기관의 HTML 소스와 구조의 분석단계를 거쳐, 인터넷상의 웹데이터를 수집·분류·저장하는 로봇을 통해 국가기관, 자치단체, 공공기관에서 제공하는 공공조달 데이터를 수집·변환하는 단계를 거쳐, 수집·변환된 다양한 패턴의 데이터(반정형 또는 비정형)는 실시간 스트림(Stream) 처리하여 RDB(Relational Database)에 적재하는 단계를 거친다. 나라장터 등 공공데이터포털에서 XML, JSON 형태로 제공하는 데이터를 수집하여 RDB에 적재하고, 한국전력공사, 수자원공사 등 HTML, 웹문서 형태로 제공하는 데이터는 전체 문서에서 기술적인 방법을 사용해 RDB에 적재한다. 비정형 데이터의 적재 방법은, 워드, 한글 문서, 이미지 등에서 데이터 수집을 하거나 파일 형태로 변경하여 RDB에 적재하며, 반정형 데이터가 아닌 대상은 자바 라이브러리를 활용하여 HTML 형태로 RDB에 적재한다.

<0044> 전처리부(330)는, 적어도 하나의 공공조달입찰 데이터가 정부입찰 집행기준 또는 정부계약 집행기준과 어긋난 오류(Error)를 찾아 수정, 정정 및 삭제 중 어느

하나 또는 적어도 하나의 조합에 대응하는 정제 및 표준화 절차를 포함하는 전처리를 수행할 수 있다. 오류가 포함된 적어도 하나의 공공조달입찰 데이터의 종류는, 예정가격범위, 예정가격, 최저가 및 예정가격에 대한 낙찰금액의 비율인 투찰율을 포함할 수 있다. 이때, 적어도 하나의 관리자 단말(400)에서 수동으로 이를 찾아서 어긋난 데이터를 맞추는 작업을 수행할 수도 있지만, 이를 자동으로 프로그램화하여 각 데이터가 서로 다른 경우, 실제 개찰 후 데이터인 확정 데이터를 기준 데이터로 하여 다른 하나의 데이터를 수정하는 방식으로 오류를 제거할 수도 있다.

<0045> 분석부(340)는, 사용자 단말(100)에서 선택한 공공입찰의 적어도 하나의 예정가격을 확률분포로 시각화하여 사용자 단말(100)로 전달할 수 있다. 사용자 단말(100)은, 공고목록에서 입찰가격을 산출하고자 하는 공고번호를 선택하고, 선택된 공고번호의 공공입찰에 참여하는 경우 낙찰가능한 적어도 하나의 예정가격(Reservation Price)을 빅데이터 기반 데이터 분석으로 추출하여 출력할 수 있다. 이때, 분석부(340)는 총 3 가지 방법을 모두 또는 적어도 하나의 조합으로 이용하여 분석을 수행할 수 있는데, 첫 번째는 빅데이터 기반, 두 번째는 데이터마이닝 기반, 세 번째는 역확률 기반이다. 이 외에도 다양한 방법으로 예측값을 도출해낼 수도 있음은 자명하다 할 것이다.

<0046> 공공입찰과 관련한 예정가격 확률분포정보 서비스 제공하는데, 공공입찰과 관련한 입찰정보, 적격심사정보, 종합심사정보 등을 제공하며, 정량적 데이터를 기반으로 빅데이터, 역확률, 데이터마이닝의 분석기술을 사용하여 공공입찰의 예정가격 확률분포 정보를 의사결정권자에게 제공함으로써 비즈니스 성과를 달성한다.

첫 번째로, 빅데이터 기반의 예정가격 확률분포 분석은, 독립변수들 간의 기술통계량을 분석하고, 분석데이터를 기반으로 상관관계를 분석하며, 상관분석을 기반으로 통계적 가설의 합당성 여부를 분석하고, 검증데이터를 기반으로 불확실한 예정가격 확률분포를 예측하여 정보를 제공한다.

<0047>

두 번째로, 역확률 기반의 예정가격 확률분포 분석은, 독립변수들 간의 기술통계량을 분석하고, 분석 데이터를 기반으로 상관관계를 분석하며, 분석 데이터를 기반으로 역확률 기반의 모델링의 분석 기술을 사용하여 예정가격 확률분포를 예측하여 정보를 제공한다. 세 번째로, 데이터마이닝 기반의 확률분포 분석 단계는, 독립변수들 간의 기술통계량을 분석하고, 분석데이터를 기반으로 상관관계를 분석하며, 분석데이터를 기반으로 데이터 마이닝 모델 등 통계적 모델링을 사용하여 예정가격 확률분포를 예측한다. 또, 의사결정권자(사용자)의 예정가격 확률분포 정보 활용할 수 있는데, 어떤 사건이 발생할 맥점가격(소수점 이하 첫째자리)을 수치로 제공하고, 어떤 사건이 발생할 복수가격(소수점 이하 둘째자리)을 수치로 제공하며, 확률분포 조건에 따라 다양한 시각적인 그래프로 제공하고, 상술한 데이터를 종합하여 최종 의사결정을 지원할 수 있다.

<0048>

빅데이터화부(350)는, 적재부(320)에 저장된 적어도 하나의 공공조달입찰 데이터인 로우 데이터(Raw Data)를 병렬 및 분산하여 저장하고, 저장된 로우 데이터 내에 포함된 비정형(Unstructured) 데이터, 정형(Structured) 데이터 및 반정형 데이터(Semi-structured)를 정제하고, 메타 데이터로 분류를 포함한 전처리를 실시하며, 전처리된 데이터를 데이터 마이닝(Data Mining)을 포함하는 분석을 실시하고,

분석된 데이터를 시각화하여 출력할 수 있다. 이때, 데이터 마이닝은, 전처리된 데이터 간의 내재된 관계를 탐색하여 클래스가 알려진 훈련 데이터 셋을 학습시켜 새로운 데이터의 클래스를 예측하는 분류(Classification) 또는 클래스 정보 없이 유사성을 기준으로 데이터를 그룹짓는 군집화(Clustering)를 수행할 수 있다.

<0049>

역확률모델부(360)는, 확률분포를 제공하기 위하여 기 설정된 적어도 하나의 독립변수 간 기술통계량을 분석한 후 상관관계 분석(Correlation Analysis)을 실행한 후, 역확률(Inverse Probability) 기반 모델링의 분석기술을 이용하여 예정가격 확률분포를 예측할 수 있다. 이때, 상관관계는 상관관계 분석(Correlation Analysis)에 의해 분석될 수 있다. 우선, 데이터 전처리 단계에서는 수집한 데이터를 사용하기 위해 데이터 형식을 변경하거나, 수집한 데이터 중에서 필요한 데이터를 추출하여 데이터의 차원을 축소할 수 있다. 특히, 자동 또는 수동으로 오류를 잡아내지 못했거나 실수로 빠트려 노이즈(Noise)가 심하고 고차원적 특성을 가지는 불필요한 독립변수로 구성되는 경우, 불필요한 데이터를 분석시스템에 포함하여 판정하면 판정 시스템의 효율을 저하시키고, 불필요한 관리비용 증가를 야기한다. 따라서, 특징 선택 방법으로 관리 데이터를 축소하면서 분석 정확도 감소를 최소화하여 품질 정책에 유의한 결과를 도출할 필요가 있다.

<0050>

특징 선택기법 중 데이터 상관관계 분석은 데이터 상관관계를 분석하여 품질 정책의 유의수준을 만족하면서 관리가 필요한 데이터를 축소하는 방법이다. 상관관계 분석에 사용되는 상관계수(Correlation Coefficient)는 두 변수 간에 어떤 선형적 관계(방향성 및 크기)가 있는지 나타낸다. 상관계수는 두 변수 간 선형관계

의 절대적 크기를 나타내는 공분산(Covariance)을 두 변수의 표준편차로 나눔으로써 방향성의 나타낸다. 두 변수를 x, y라 할 때, 피어슨(Pearson) 상관계수를 수식으로 나타내면 이하 수학적 식 1과 같다.

<0051>

【수학적 식 1】

$$r = r_{xy} = \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

<0052>

상관 관계분석 단계에서는 피어슨 상관분석을 통하여 독립변수들의 상관계수를 도출하고, 데이터 연관성이 큰 유효 인자를 선별하여 중복 특징을 가지는 데이터를 제거할 수 있다. 그리고, SVM(Support Vector Machine) 매개변수 최적화 단계에서는 Cross Validation 기반으로 SVM 매개변수 값을 변경하면서 매개변수의 최적값을 찾을 수 있다. SVM 테스트 및 정확도 분석 단계에서는 커널에 따른 성능을 비교 분석하여 정확도가 가장 높은 커널을 선정할 수 있다. 또 가장 높은 정확도를 가지는 커널과 SVM 매개변수를 시스템에 적용할 수 있다. 이때, 상관계수를 이용하여 최적 특징 선택을 찾는 Fast Correlation Based Filter Approach 알고리즘을 이용할 수 있으나 이에 한정되는 것은 아니다. 또, 피어슨 상관계수는 1에 가까울수록 양의 상관관계를 나타내고, -1에 가까울수록 음의 상관관계를 나타낸다. 만약 피어슨 상관계수가 0에 가깝다면, 상관관계가 희박함을 의미한다. 상관계수의 유의 확률(P-value)이 설정된 유의수준(Significant Level) 이하일 때, 상관계수가 유의하다고 할 수 있다.

<0053> 이러한 상관관계 분석은 빅데이터 모델링, 데이터마이닝 모델링 및 역확률 모델링에 모두 사용될 수 있다. 또, 상관관계 분석을 위한 방법이 피어슨 이외에도 다양한 방법이 사용될 수도 있다.

<0054> 이하, 상술한 도 2의 예측 서비스 제공 서버의 구성에 따른 동작 과정을 도 3 및 도 4를 예로 들어 상세히 설명하기로 한다. 다만, 실시예는 본 발명의 다양한 실시예 중 어느 하나일 뿐, 이에 한정되지 않음은 자명하다 할 것이다.

<0055> 도 3을 참조하면, 본 발명의 일 실시예에 따른 서비스(비드프로)는 도 3과 같은 플로우로 본 발명의 일 실시예에 따른 서비스를 진행할 수 있다. 이때, 비드 프로는 본 발명의 일 실시예에 따른 예측 서비스 제공 서버(300), 의사결정권자는 사용자 단말(100)의 사용자(User)이다. 본 발명의 일 실시예에 따른 예측 서비스 제공 서버(300)는 공공조달 자료를 수집하여 검수를 한 후 정량적 정제를 수행하고 적재를 한다. 이때, 오류 자료는 상술한 바와 같이 수동 또는 자동으로 검수할 수 있으며, 오류가 정제되고 표준화된 데이터는 데이터의 종류에 따라 정형, 반정형 및 비정형에 맞게 DB가 정해져 적재된다. 또, 세 가지의 모형, 즉 빅데이터 모형, 역확률 모형 및 데이터 마이닝 모형 중 어느 하나 또는 적어도 하나의 조합으로 확률분포를 예측하는데, 통계적 모델링 적용방식이면 바로 확률분포 정보를 제공하고, 이때, 예가범위(예정가격범위)를 자동산정하는 경우 랜덤분포 또는 정규분포에 따른 확률분포를 제공한다. 또, 복수예비가격 적용구분이 예가공개 또는 비공개인 경우에 따라 다시 공개 예측과 비공개 예측을 수행하여 최종 예정가격을 제공한다.

<0056> <공고목록에서 입찰가격을 산출하고자 하는 공고번호를 클릭>

<0057> 도 4a를 참조하면, 사용자 단말(100)은, 예를 들어, 2021.02.16. 마감 되는 공고 중 관리를 체크한 3건의 입찰가격을 산출할 수 있다. 사용자 단말(100)은, 로그인 후 [입찰정보]를 클릭하면 입찰정보시스템이 실행된다. 공고목록에서 입찰가격을 산출하고자 하는 공고번호를 클릭하면 각호의 정보를 보여준다(관리체크 공고번호: 20201242334, 20210126177, 20210126191). ① 입찰참가자격 등 입찰공고 정보를 보여주고, ② 입찰가격 산출과 관련된 정보(기초금액, 투찰률, 순공사원가 등)를 보여주며, ③ 적격심사점수 또는 종합심사점수를 보여준다. 이때, (-X)는, 부적격(수행점수가 X점 부족하다는 의미), 생략은, 적격심사 생략(입찰가격만으로 최종낙찰자를 확정한다는 의미), 공동은, 공동계약(공동수급체의 합산 수행능력점수를 산출하라는 의미), 통과는, 입찰참가자격을 모두 충족하였으므로 단독으로 입찰참가를 할 수 있다는 의미이다.

<0058> <맥+차트 클릭, 매뉴얼을 실행하여 낙찰지점을 예측>

<0059> 도 4b를 참조하면, 사용자 단말(100)은, [맥+차트] 버튼을 클릭하여 예정가격 예측시스템을 실행한다. 예정가격 예측시스템은 복수예비가격 조합 비율 중 낙찰구간과 조합 가능한 최적 숫자를 예측하여 낙찰지점 1개를 제시하는 도구이다. 사용자 단말(100)은, 매뉴얼을 실행하여 낙찰지점을 예측하며 각 기능은 이하 표 1과 같다.

<0060> 【표 1】

구분	위치	매뉴얼 실행	매뉴얼 실행결과
원칙	1	4개 구간 중 최고확률 구간 클릭	최고확률 낙찰구간 제시
	2	맥점 클릭	2~3개 낙찰지점 제시
	3	필터 클릭	1~2개 낙찰지점 제시

	4	맥점(↓↑) 중 녹색표시가 된 맥점	1개 낙찰지점 제시
최종결과	6	광고번호 20201242334 실행	99.2% 예측(낙찰지점)

<0061> <낙찰지점을 클릭, 매뉴얼을 실행하여 최종 입찰가격을 산출>

<0062> 도 4c를 참조하면, 사용자 단말(100)은, 낙찰지점(99.2%)을 클릭, 복수예가 예측시스템을 실행한다. 복수예가 예측시스템은, 낙찰지점에 해당하는 100개의 입찰가격 중 낙찰지점과 조합 가능한 최적 숫자를 예측하여 최종 입찰가격 1개를 제시하는 도구이다. 사용자 단말(100)은, 매뉴얼을 실행하여 최종 입찰가격을 산출하며 이하 표 2와 같다.

<0063> 【표 2】

구분	위치	매뉴얼 실행	매뉴얼 실행결과
소수점 이하 둘째 자리 수 예측	1	저점과 고점 중 최고 확률숫자 클릭	최고확률 숫자 2개 제시
	2		
소수점 이하 셋째 자리 수 예측	-	소수점이하 둘째자리 분석과 동일	셋째자리 숫자 8 제시
최종결과	3	광고번호 20201242334 실행	₩ 25,335,897,451원 산출(99.2789%)

<0064> <최종 확정된 입찰가격을 클릭하여 저장>

<0065> 도 4d를 참조하면, 사용자 단말(100)은, 매뉴얼 실행을 모두 완료하면 공고 목록으로 복귀하고, 최종 확정된 입찰가격을 투찰목록으로 송신한다. 최종 확정된 입찰가격은 낙찰자 선정에서 제외되지 않도록 순공사원가의 98%이상과 법정보험료(A값)가 적용되어 있다. ① 최종 입찰가격을 클릭하면 입찰가격을 산출한 근거를 보여주고, ② [저장] 버튼을 클릭하면 입찰가격이 임시 저장되고, ③ [송신]을 체크하면 입찰가격이 투찰목록에 저장되고, ④는 투찰목록을 클릭하면 저장한 입찰가격을 확인할 수 있고, ⑤ 적격심사제의 입찰에서 자사의 적격점수가 미달되면 부족점수 보완항목의 입력박스에 부족한 점수를 입력하고 ‘가격보완이전’을 클릭한

다.

<0066> 이와 같은 도 2 내지 도 4의 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법에 대해서 설명되지 아니한 사항은 앞서 도 1을 통해 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법에 대하여 설명된 내용과 동일하거나 설명된 내용으로부터 용이하게 유추 가능하므로 이하 설명을 생략하도록 한다.

<0067> 도 5는 본 발명의 일 실시예에 따른 도 1의 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템에 포함된 각 구성들 상호간에 데이터가 송수신되는 과정을 나타낸 도면이다. 이하, 도 5를 통해 각 구성들 상호간에 데이터가 송수신되는 과정의 일 예를 설명할 것이나, 이와 같은 실시예로 본원이 한정 해석되는 것은 아니며, 앞서 설명한 다양한 실시예들에 따라 도 5에 도시된 데이터가 송수신되는 과정이 변경될 수 있음은 기술분야에 속하는 당업자에게 자명하다.

<0068> 도 5를 참조하면, 예측 서비스 제공 서버는, 적어도 하나의 공공조달입찰 데이터를 웹크롤러를 이용하여 수집하고(S5100), 수집된 적어도 하나의 공공조달입찰 데이터 내 정형(Structured) 데이터, 반정형(Semi-Structured) 및 비정형(Unstructured) 데이터를 적재한다(S5200).

<0069> 또, 예측 서비스 제공 서버는, 적어도 하나의 공공조달입찰 데이터가 정부입찰 집행기준 또는 정부계약 집행기준과 어긋난 오류(Error)를 찾아 수정, 정정 및 삭제 중 어느 하나 또는 적어도 하나의 조합에 대응하는 정제 및 표준화 절차를 포

함하는 전처리를 수행하고(S5300), 사용자 단말에서 선택한 공공입찰의 적어도 하나의 예정가격을 확률분포로 시각화하여 사용자 단말로 전달한다(S5400).

<0070> 상술한 단계들(S5100~S5400)간의 순서는 예시일 뿐, 이에 한정되지 않는다. 즉, 상술한 단계들(S5100~S5400)간의 순서는 상호 변동될 수 있으며, 이중 일부 단계들은 동시에 실행되거나 삭제될 수도 있다.

<0071> 이와 같은 도 5의 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법에 대해서 설명되지 아니한 사항은 앞서 도 1 내지 도 3을 통해 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법에 대하여 설명된 내용과 동일하거나 설명된 내용으로부터 용이하게 유추 가능하므로 이하 설명을 생략하도록 한다.

<0072> 도 5를 통해 설명된 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법은, 컴퓨터에 의해 실행되는 애플리케이션이나 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

<0073> 전술한 본 발명의 일 실시예에 따른 공공조달시장의 입찰을 위한 빅데이터

기반 예정가격 확률분포 예측 서비스 제공 방법은, 단말기에 기본적으로 설치된 애플리케이션(이는 단말기에 기본적으로 탑재된 플랫폼이나 운영체제 등에 포함된 프로그램을 포함할 수 있음)에 의해 실행될 수 있고, 사용자가 애플리케이션 스토어 서버, 애플리케이션 또는 해당 서비스와 관련된 웹 서버 등의 애플리케이션 제공 서버를 통해 마스터 단말기에 직접 설치한 애플리케이션(즉, 프로그램)에 의해 실행될 수도 있다. 이러한 의미에서, 전술한 본 발명의 일 실시예에 따른 공공조달 시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 방법은 단말기에 기본적으로 설치되거나 사용자에게 의해 직접 설치된 애플리케이션(즉, 프로그램)으로 구현되고 단말기에 등의 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다.

<0074> 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

<0075> 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야

한다.

## 【청구범위】

### 【청구항 4】

공고목록에서 입찰가격을 산출하고자 하는 공고번호를 선택하고, 선택된 공고번호의 공공입찰에 참여하는 경우 낙찰가능한 적어도 하나의 예정가격(Reservation Price)을 빅데이터 기반 데이터 분석으로 추출하여 출력하는 사용자 단말; 및

적어도 하나의 공공조달입찰 데이터를 웹크롤러를 이용하여 수집하는 수집부, 상기 수집된 적어도 하나의 공공조달입찰 데이터 내 정형(Structured) 데이터, 반정형(Semi-Structured) 및 비정형(Unstructured) 데이터를 적재하는 적재부, 상기 적어도 하나의 공공조달입찰 데이터가 정부입찰 집행기준 또는 정부계약 집행기준과 어긋난 오류(Error)를 찾아 수정, 정정 및 삭제 중 어느 하나 또는 적어도 하나의 조합에 대응하는 정제 및 표준화 절차를 포함하는 전처리를 수행하는 전처리부, 상기 사용자 단말에서 선택한 공공입찰의 적어도 하나의 예정가격을 확률분포로 시각화하여 상기 사용자 단말로 전달하는 분석부를 포함하는 예측 서비스 제공 서버;

를 포함하되,

상기 적재부는,

상기 정형 데이터는 RDBMS(Relational DataBase Management System) 데이터베이스 저장방식으로 MSSQL(Microsoft SQL)에 저장하고,

상기 반정형 데이터는 RDB(Relational Database)에 저장하고,

상기 비정형 데이터는, 분산파일시스템인 DFS(Distributed File System)에 저장하되,

상기 적어도 하나의 공공조달입찰 데이터를 실시간 스트림(Realtime Stream) 처리하여 적재하고,

상기 오류가 포함된 적어도 하나의 공공조달입찰 데이터의 종류는,

예정가격범위, 예정가격, 최저가 및 예정가격에 대한 낙찰금액의 비율인 투찰율을 포함하는 것을 특징으로 하는 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템.

#### 【청구항 5】

제 4 항에 있어서,

상기 예측 서비스 제공 서버는,

상기 적재부에 저장된 적어도 하나의 공공조달입찰 데이터인 로우 데이터(Raw Data)를 병렬 및 분산하여 저장하고, 상기 저장된 로우 데이터 내에 포함된 비정형(Unstructured) 데이터, 정형(Structured) 데이터 및 반정형 데이터(Semi-structured)를 정제하고, 메타 데이터로 분류를 포함한 전처리를 실시하며, 상기 전처리된 데이터를 데이터 마이닝(Data Mining)을 포함하는 분석을 실시하고, 상기 분석된 데이터를 시각화하여 출력하는 빅데이터화부;

를 더 포함하는 것을 특징으로 하는 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템.

**【청구항 6】**

제 5 항에 있어서,

상기 데이터 마이닝은,

상기 전처리된 데이터 간의 내재된 관계를 탐색하여 클래스가 알려진 훈련 데이터 셋을 학습시켜 새로운 데이터의 클래스를 예측하는 분류(Classification) 또는 클래스 정보 없이 유사성을 기준으로 데이터를 그룹짓는 군집화(Clustering)를 수행하는 것을 포함하는 것을 특징으로 하는 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템.

**【청구항 7】**

제 4 항에 있어서,

상기 예측 서비스 제공 서버는,

상기 확률분포를 제공하기 위하여 기 설정된 적어도 하나의 독립변수 간 기술통계량을 분석한 후 상관관계 분석(Correlation Analysis)을 실행한 후, 역확률(Inverse Probability) 기반 모델링의 분석기술을 이용하여 예정가격 확률분포를 예측하는 역확률모델부;

를 더 포함하는 것을 특징으로 하는 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템.

## 【요약서】

### 【요약】

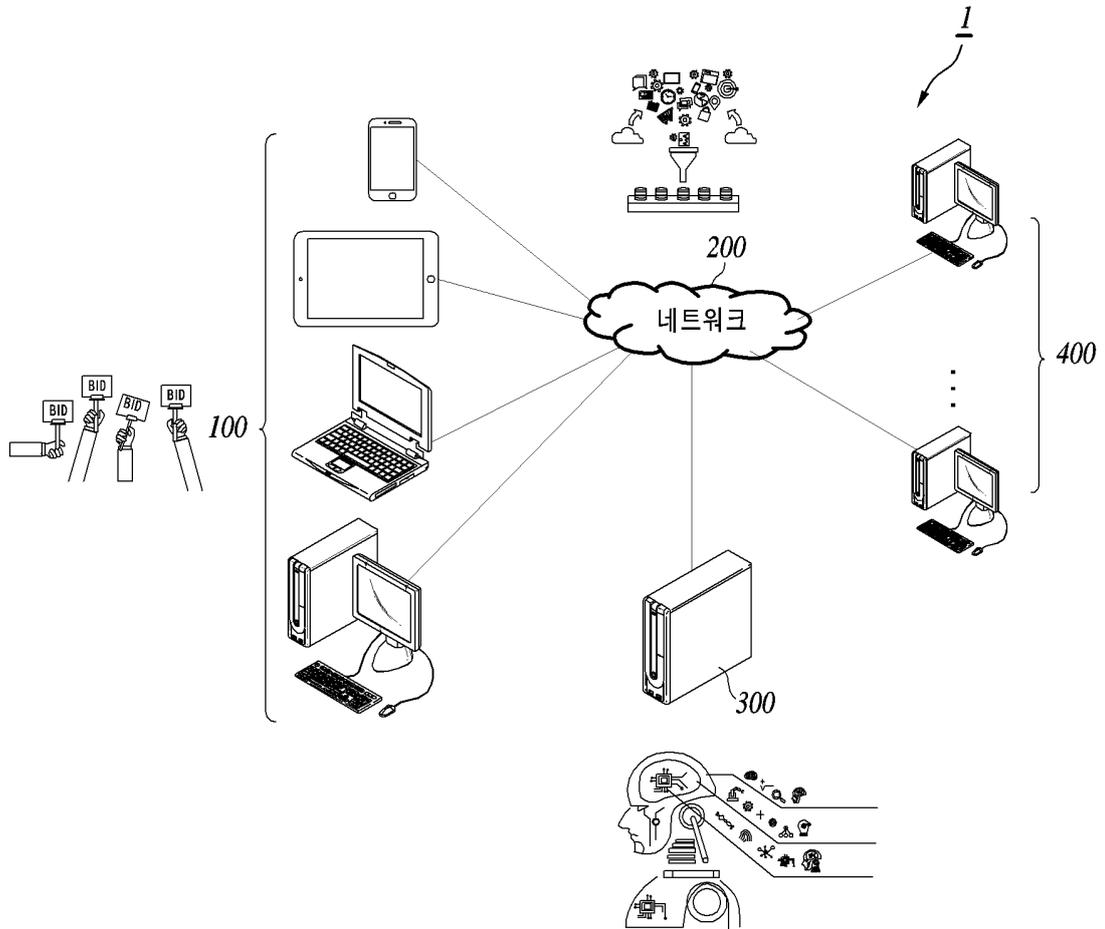
공공조달시장의 입찰을 위한 빅데이터 기반 예정가격 확률분포 예측 서비스 제공 시스템이 제공되며, 공고목록에서 입찰가격을 산출하고자 하는 공고번호를 선택하고, 선택된 공고번호의 공공입찰에 참여하는 경우 낙찰가능한 적어도 하나의 예정가격(Reservation Price)을 빅데이터 기반 데이터 분석으로 추출하여 출력하는 사용자 단말 및 적어도 하나의 공공조달입찰 데이터를 웹크롤러를 이용하여 수집하는 수집부, 수집된 적어도 하나의 공공조달입찰 데이터 내 정형(Structured) 데이터, 반정형(Semi-Structured) 및 비정형(Unstructured) 데이터를 적재하는 적재부, 적어도 하나의 공공조달입찰 데이터가 정부입찰 집행기준 또는 정부계약 집행기준과 어긋난 오류(Error)를 찾아 수정, 정정 및 삭제 중 어느 하나 또는 적어도 하나의 조합에 대응하는 정제 및 표준화 절차를 포함하는 전처리를 수행하는 전처리부, 사용자 단말에서 선택한 공공입찰의 적어도 하나의 예정가격을 확률분포로 시각화하여 사용자 단말로 전달하는 분석부를 포함하는 예측 서비스 제공 서버를 포함한다.

### 【대표도】

도 1

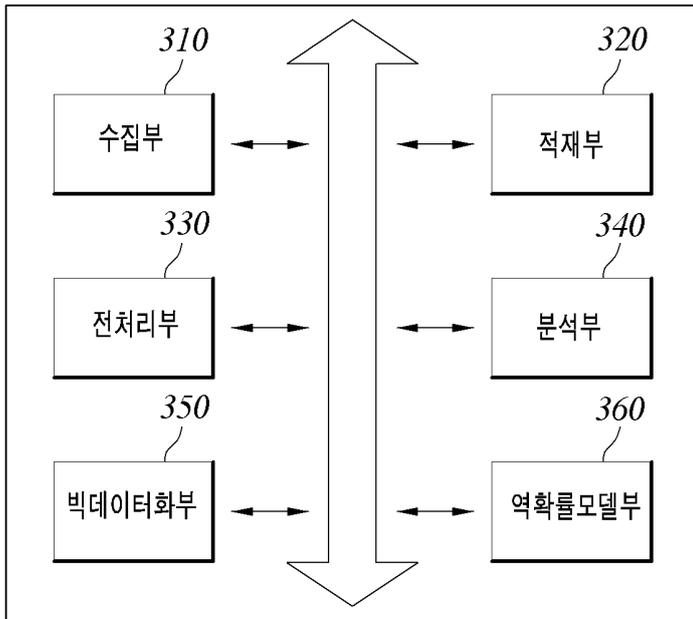
【도면】

【도 1】

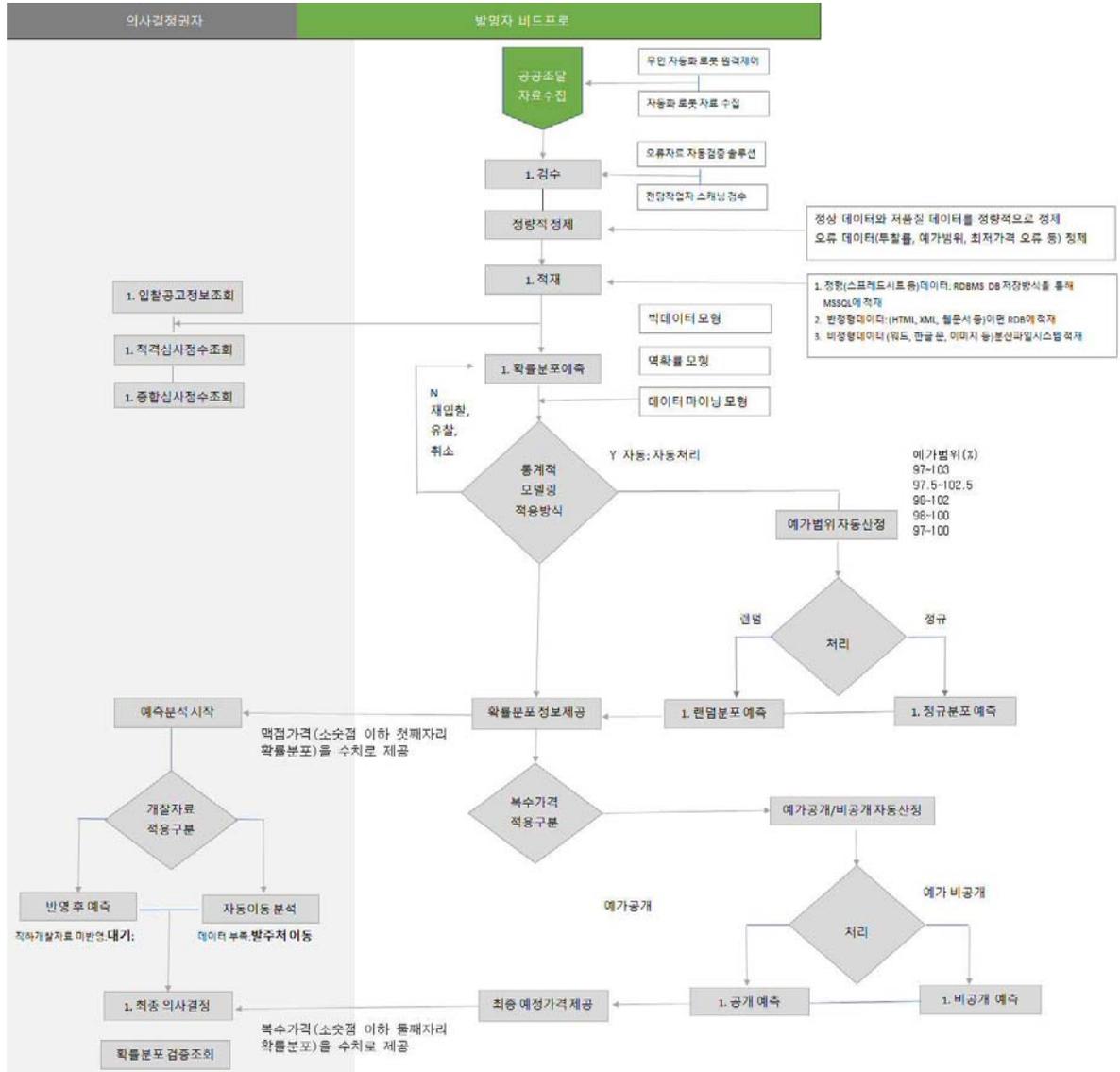


【도 2】

300



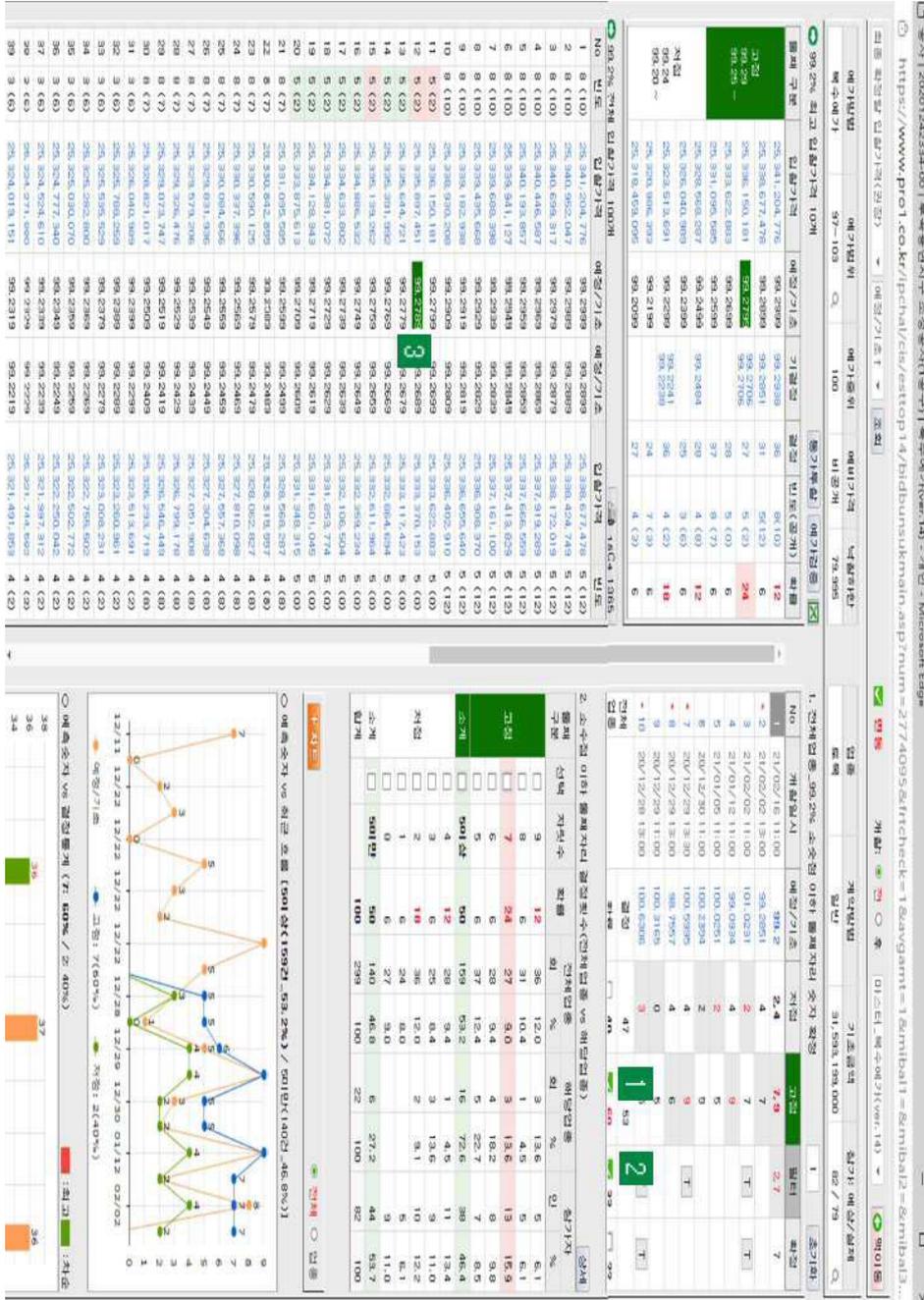
【도 3】







【도 4c】





【도 5】

