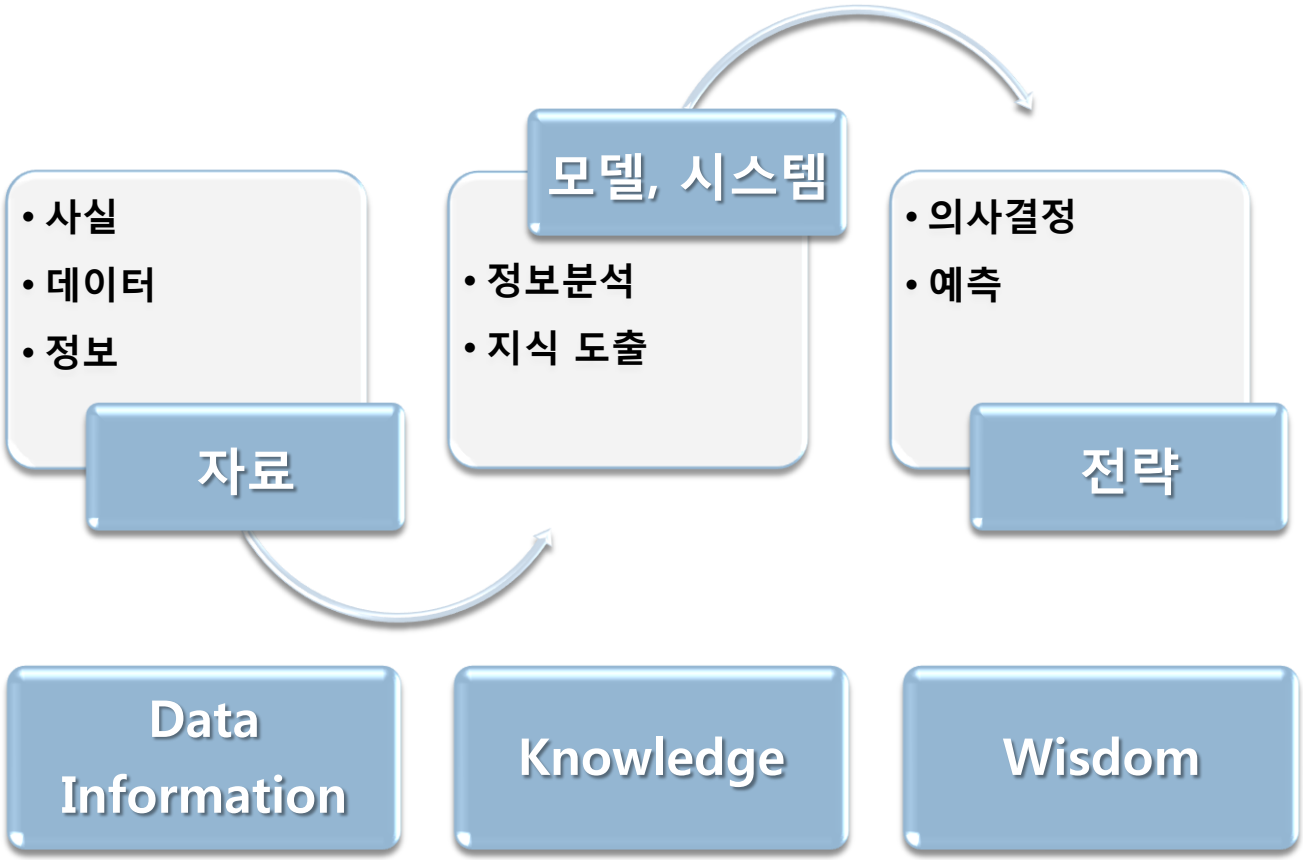
빅데이터 방법론 기반 연구사례

2014. 09. 29

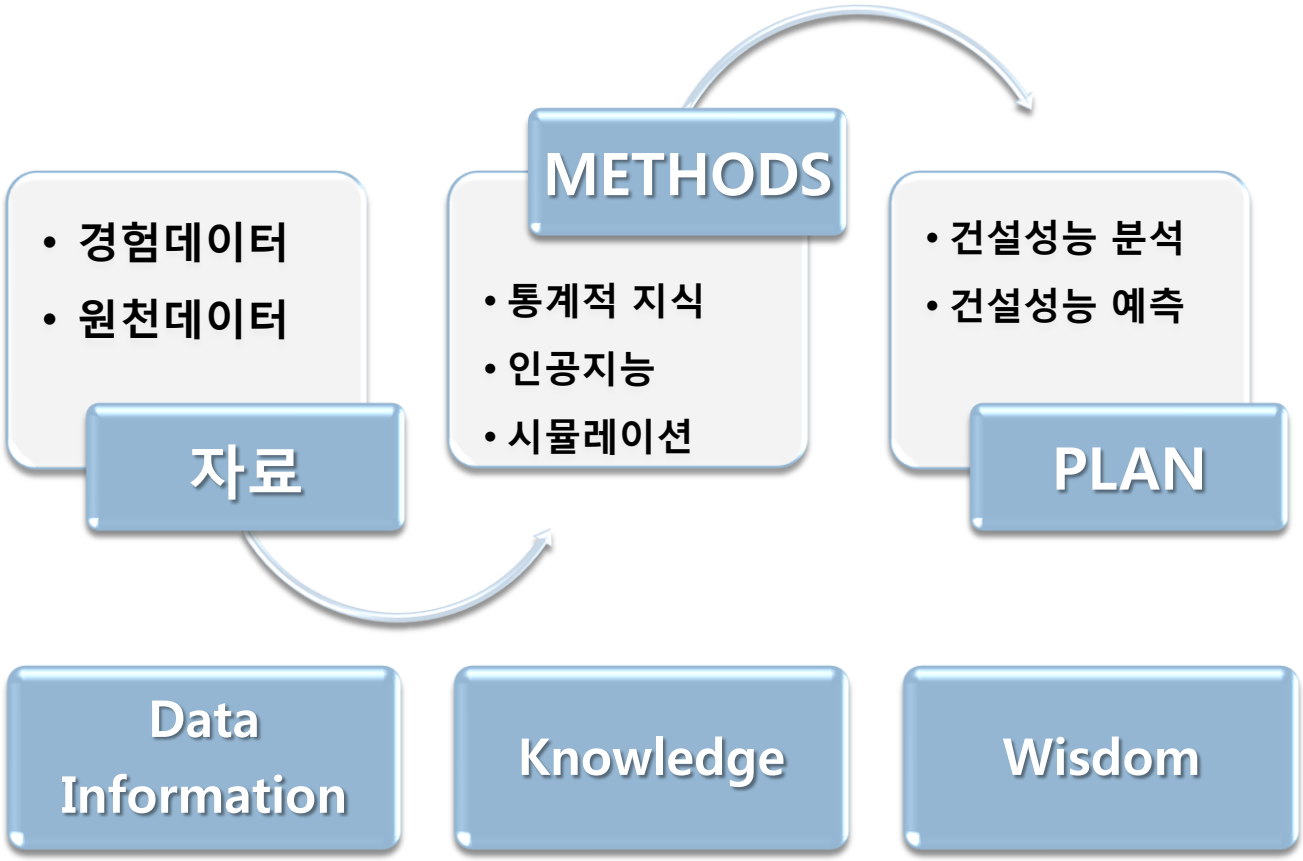
한 승 우
인하대학교 공과대학 건축공학과

인하대학교

개요 1. 데이터기반 지식시스템



개요 2. 건설분야의 지식시스템



개요 2. 건설분야의 지식시스템

METHODS

- 통계적 지식
- 인공지능
- 시뮬레이션

- 건설성능 분석
- 건설성능 예측

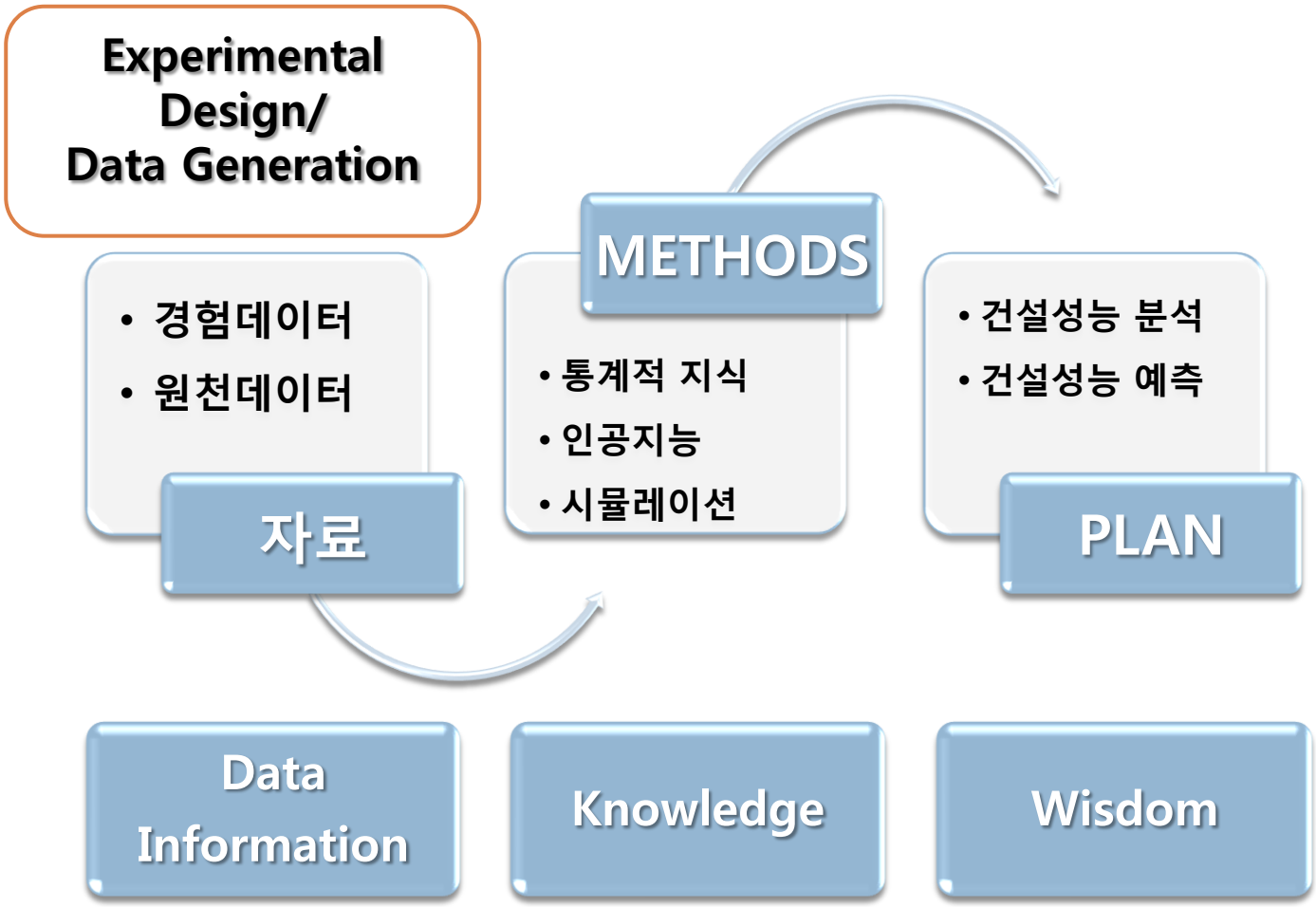
PLAN

Data
Information

Knowledge

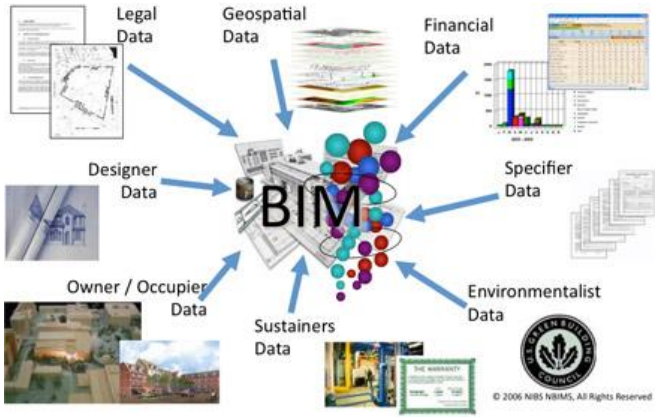
Wisdom

개요 2. 건설분야의 지식시스템

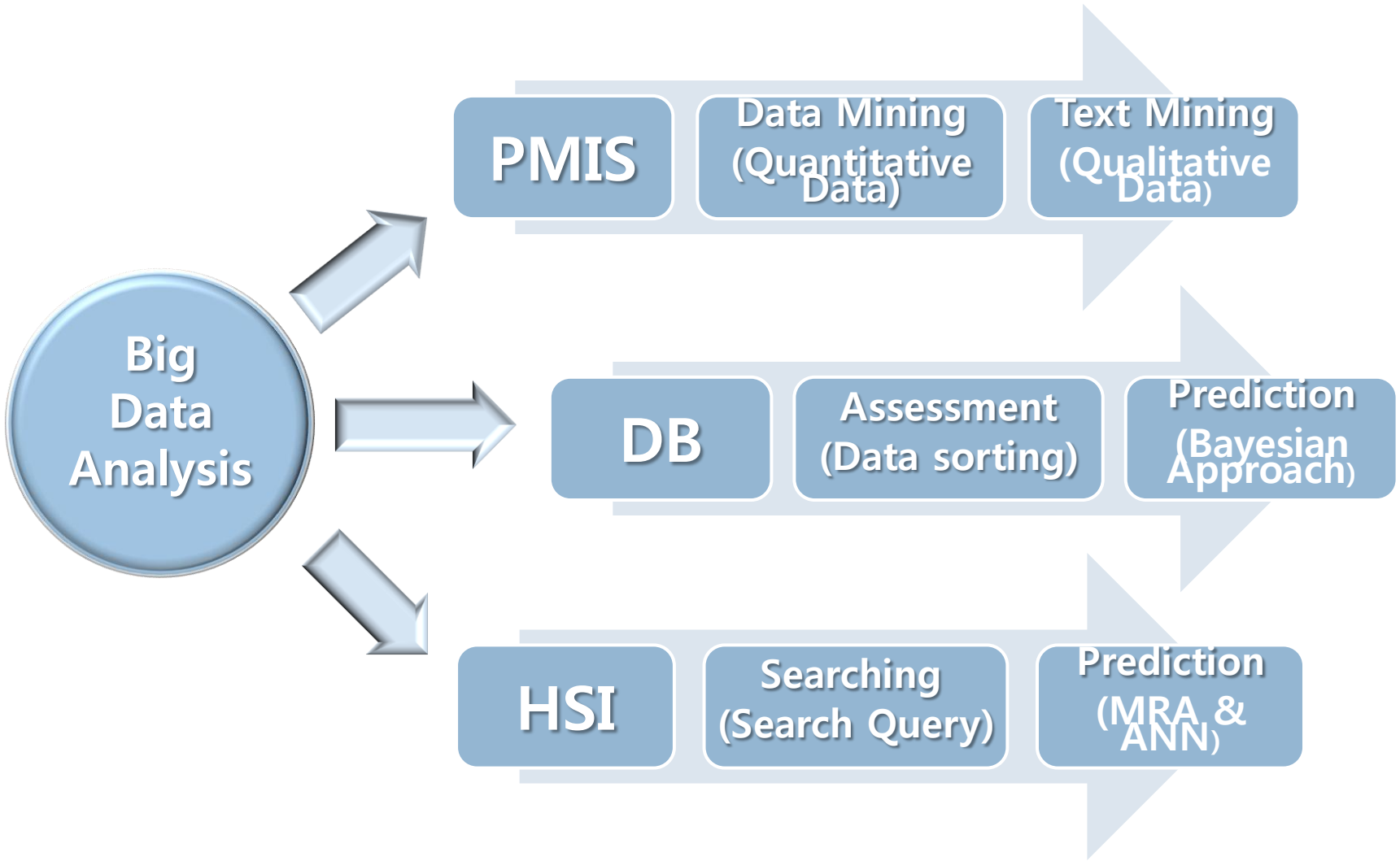


빅데이터 분석기반 연구 (수행)

건설원천 빅데이터 속성 분석모형 개발
한국연구재단 중견연구자지원사업(핵심연구: 개인)
2012.12~2015.11



빅데이터 개념 기반 연구과제 (수행)



TEXT MINING ON PMIS

Actual Data Assessment

열1	하부보기	하역작업	하차	하현재	한국지질	할석	할암	함설치	합벽	합벽거푸집	합벽부위	합벽비닐	합벽앙카	합벽지지대	합벽형틀	합판	해체	해체작업
2012년 2월 21일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4
2012년 2월 22일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2012년 2월 23일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
2012년 2월 24일	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0
2012년 2월 25일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1
2012년 2월 26일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
2012년 2월 27일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
2012년 2월 28일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
2012년 2월 29일	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0

average	0.00	0.05	0.03	0.01	0.02	0.05	0.03	0.01	0.16	0.14	0.01	0.01	0.01	0.01	0.01	0.10	0.17	0.31
sum	1	16	9	4	7	17	9	3	54	47	3	2	2	4	4	32	56	105
5						17	9								*	32	56	105
10						17	*								*	32	56	105
20						*	*								*	32	56	105
50						*	*								*	*	56	105
100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	105
200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
300	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Daily words counting.

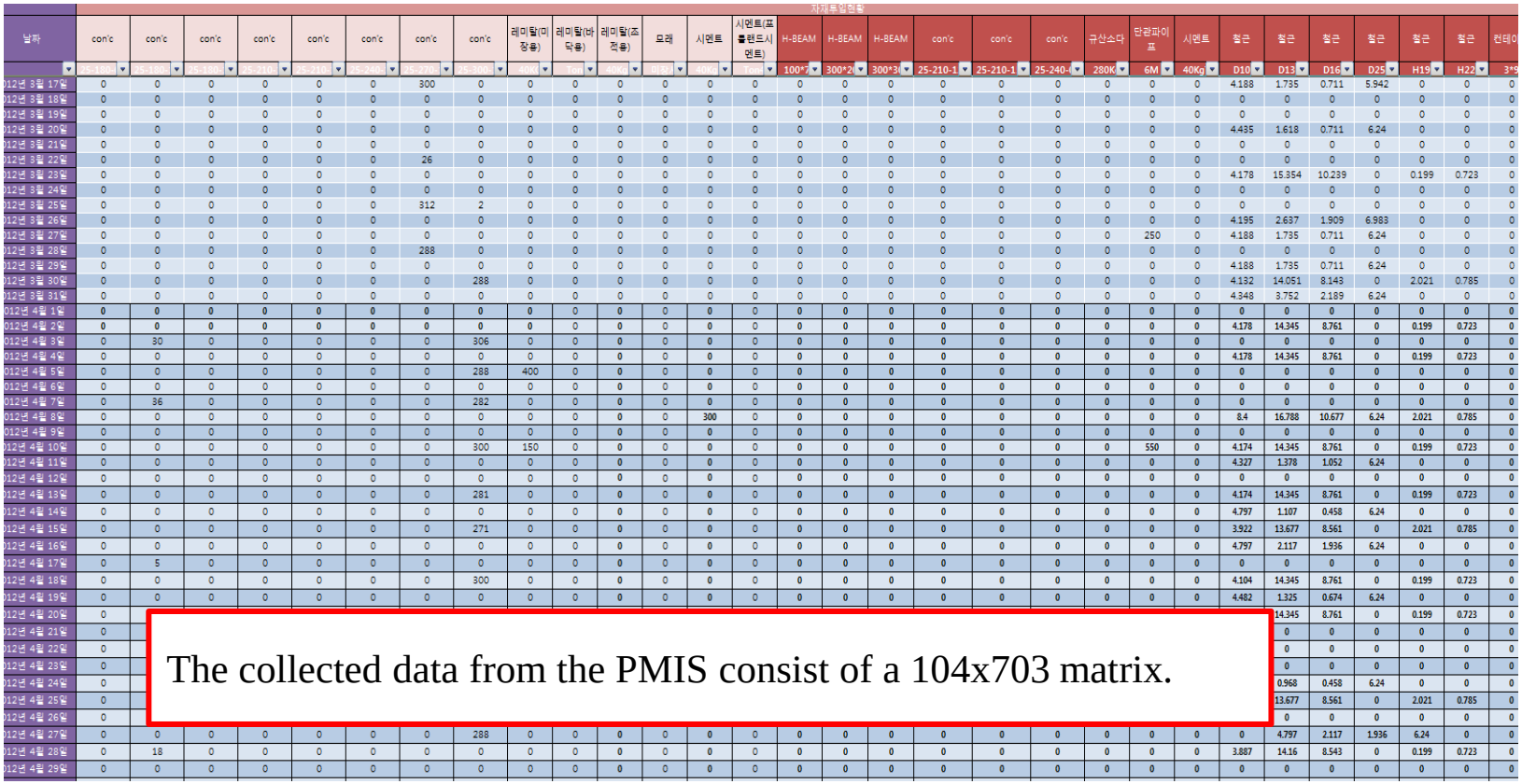
Calculation of average and sum.

TEXT MINING ON PMIS

Actual Data Assessment

104

703



날짜	con'c	con'c	con'c	con'c	con'c	con'c	con'c	con'c	레미탈(미장용)	레미탈(바닥용)	레미탈(조적용)	모래	시멘트	시멘트(포틀랜드시멘트)	H-BEAM	H-BEAM	H-BEAM	con'c	con'c	con'c	규산소다	단관파이프	시멘트	철근	철근	철근	철근	철근	철근	컨테이
2012년 3월 17일	0	0	0	0	0	0	300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.188	1.735	0.711	5.942	0	0	0
2012년 3월 18일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 19일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 20일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.435	1.618	0.711	6.24	0	0	0
2012년 3월 21일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 22일	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 23일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.178	15.354	10.239	0	0.199	0.723	0
2012년 3월 24일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 25일	0	0	0	0	0	0	312	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 26일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.195	2.637	1.909	6.983	0	0	0
2012년 3월 27일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	250	0	4.188	1.735	0.711	6.24	0	0	0
2012년 3월 28일	0	0	0	0	0	0	288	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 3월 29일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.188	1.735	0.711	6.24	0	0	0
2012년 3월 30일	0	0	0	0	0	0	0	288	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.132	14.051	8.143	0	2.021	0.785	0
2012년 3월 31일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.348	3.752	2.189	6.24	0	0	0
2012년 4월 1일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 2일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.178	14.345	8.761	0	0.199	0.723	0
2012년 4월 3일	0	30	0	0	0	0	0	306	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 4일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.178	14.345	8.761	0	0.199	0.723	0
2012년 4월 5일	0	0	0	0	0	0	0	288	400	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 6일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 7일	0	36	0	0	0	0	0	282	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 8일	0	0	0	0	0	0	0	0	0	0	0	0	300	0	0	0	0	0	0	0	0	0	0	8.4	16.788	10.677	6.24	2.021	0.785	0
2012년 4월 9일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 10일	0	0	0	0	0	0	0	300	150	0	0	0	0	0	0	0	0	0	0	0	0	550	0	4.174	14.345	8.761	0	0.199	0.723	0
2012년 4월 11일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.327	1.378	1.052	6.24	0	0	0
2012년 4월 12일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 13일	0	0	0	0	0	0	0	281	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.174	14.345	8.761	0	0.199	0.723	0
2012년 4월 14일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.797	1.107	0.458	6.24	0	0	0
2012년 4월 15일	0	0	0	0	0	0	0	271	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.922	13.677	8.561	0	2.021	0.785	0
2012년 4월 16일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.797	2.117	1.936	6.24	0	0	0
2012년 4월 17일	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012년 4월 18일	0	0	0	0	0	0	0	300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.104	14.345	8.761	0	0.199	0.723	0
2012년 4월 19일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.482	1.325	0.674	6.24	0	0	0
2012년 4월 20일	0																								14.345	8.761	0	0.199	0.723	0
2012년 4월 21일	0																								0	0	0	0	0	0
2012년 4월 22일	0																								0	0	0	0	0	0
2012년 4월 23일	0																								0	0	0	0	0	0
2012년 4월 24일	0																								0.968	0.458	6.24	0	0	0
2012년 4월 25일	0																								13.677	8.561	0	2.021	0.785	0
2012년 4월 26일	0																								0	0	0	0	0	0
2012년 4월 27일	0	0	0	0	0	0	0	288	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.797	2.117	1.936	6.24	0	0
2012년 4월 28일	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.887	14.16	8.543	0	0.199	0.723	0
2012년 4월 29일	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The collected data from the PMIS consist of a 104x703 matrix.

TEXT MINING ON PMIS

Monthly and Weekly Basis Data Summation

	마감	바닥	배근작업	배근	배근sum	벽체	설치	설치sum	슬라브	자재	자재sum	미장	미장sum	방수	방수sum	작업	작업sum
2011년 4월 30일	0	0	0	0	0	0	6	6	0	0	1	0	0	0	0	0	4
2011년 5월 31일	0	0	0	0	0	0	8	34	0	0	0	0	0	0	0	2	24
2011년 6월 30일	0	0	0	0	0	0	110	128	0	3	3	0	0	0	0	36	62
2011년 7월 31일	1	0	0	0	0	0	136	158	0	0	0	0	0	0	0	37	94
2011년 8월 31일	0	1	0	0	0	0	179	213	0	2	2	0	0	0	0	9	56
2011년 9월 30일	2	19	0	0	0	0	19	19	0	4	4	0	0	0	0	42	65
2011년 10월 31일	16	10	11	0	11	14	63	65	50	4	5	2	3	0	0	121	191
2011년 11월 30일	12	20	13	0	13	8	31	34	44	37	37	2	3	0	0	128	226
2011년 12월 31일	11	47	17	0	17	27	60	60	46	25	25	5	5	0	0	206	272
2012년 1월 31일	24	66	17	0	17	20	91	99	42	13	13	1	1	0	0	247	336
2012년 2월 29일	13	88	11	0	11	29	108	110	47	13	13	14	14	0	0	263	394
2012년 3월 31일	5	99	0	23	23	25	77	77	40	36	37	35	35	5	22	300	348
2012년 4월 30일	20	120	0	39	39	28	103	103	23	19	19	61	61	0	31	347	382
2012년 5월 31일	19	66	0	26	26	12	161	161	5	20	20	77	77	12	53	408	446
2012년 6월 17일	4	7	0	8	8	13	85	85	1	11	17	45	45	5	13	247	275
2012년 6월 30일	0	0	15	1	16	28	64	103	0	14	21	0	34	0	20	0	148
2012년 7월 31일	0	1	2	0	2	75	189	247	0	53	86	2	75	13	41	10	278
2012년 8월 24일	6	5	1	4	5	33	125	209	0	27	50	4	64	13	15	41	208
2012년 8월 31일	0	6	0	1	1	5	41	105	0	5	5	10	10	11	12	119	128
2012년 9월 30일	6	17	0	2	2	38	158	283	0	41	41	23	23	32	36	304	348
2012년 10월 31일	27	29	0	3	3	101	286	310	0	26	26	17	17	42	43	531	562
2012년 11월 10일	5	18	0	0	0	23	92	95	0	14	14	4	4	11	11	153	159
2012년 11월 30일	8	35	0	0	0	72	258	266	0	6	13	15	15	12	15	464	491
2012년 12월 31일	2	79	0	0	0	67	340	349	0	16	19	10	10	2	6	497	534
2013년 1월 31일	1	7	0	0	0	15	155	163	0	11	13	7	7	0	1	247	268
2013년 2월 28일	0	3	0	0	0	0	14	14	0	1	1	1	1	0	0	26	32
	182	743	87	107	194	633	2959	2959	298	401	401	335	335	158	158	4785	##

TEXT MINING ON PMIS

Monthly and Weekly Basis Data Summation

	con'c 25-180-08	con'c 25-180-12	con'c 25-180-15	con'c 25-210-12	con'c 25-210-15	con'c 25-240-15	con'c 25-270-15	con'c 25-300-15	con'c 25-210-125	con'c 25-210-156	con'c 25-240-08	conc sum	미
2011년 4월 30일	14	0	0	42	18	0	0	0	0	0	0	74	
2011년 5월 31일	0	0	0	167	251	0	0	0	0	0	0	418	
2011년 6월 30일	0	0	0	0	15	0	0	0	0	0	0	15	
2011년 7월 31일	0	0	0	0	0	0	0	0	0	0	0	0	
2011년 8월 31일	0	0	0	0	0	0	0	0	0	0	0	0	
2011년 9월 30일	0	0	510	0	0	2112	0	0	0	0	0	2622	
2011년 10월 31일	0	0	0	0	0	1953	2884	0	0	0	0	4837	
2011년 11월 30일	0	0	0	0	0	0	2579	0	0	0	0	2579	
2011년 12월 31일	0	0	0	0	0	0	2863	0	0	0	0	2863	
2012년 1월 31일	0	0	0	0	0	0	2409	0	0	0	0	2409	
2012년 2월 29일	0	0	0	0	0	0	1300	0	0	0	0	1300	
2012년 3월 31일	0	0	0	0	0	0	1987	290	200	0	0	2477	
2012년 4월 30일	0	125	0	0	0	0	0	2605	1350	0	0	4080	
2012년 5월 31일	0	347	0	0	0	0	1114	1285	1600	0	0	4346	
2012년 6월 30일	0	59	0	0	0	0	562	0	5900	0	0	6521	
2012년 7월 31일	0	97	2	0	0	0	0	0	1050	0	0	1149	
2012년 8월 31일	0	304	24	3	0	2	4	103	7961	706.7	2	9109.7	
2012년 9월 30일	0	337	0	1	0	3	1	338	0	0	3	683	
2012년 10월 31일	0	279	0	0	3	5	14	104	1851	0	0	2256	
2012년 11월 30일	0	131	0	0	0	0	0	0	2877	444.7	0	3452.7	
2012년 12월 31일	0	80	0	2	0	0	0	0	0	0	0	82	
2013년 1월 31일	0	0	0	0	0	0	0	0	0	0	0	0	
2013년 2월 28일	0	0	0	0	0	0	0	0	0	0	0	0	

TEXT MINING ON PMIS

Correlation Analysis using R

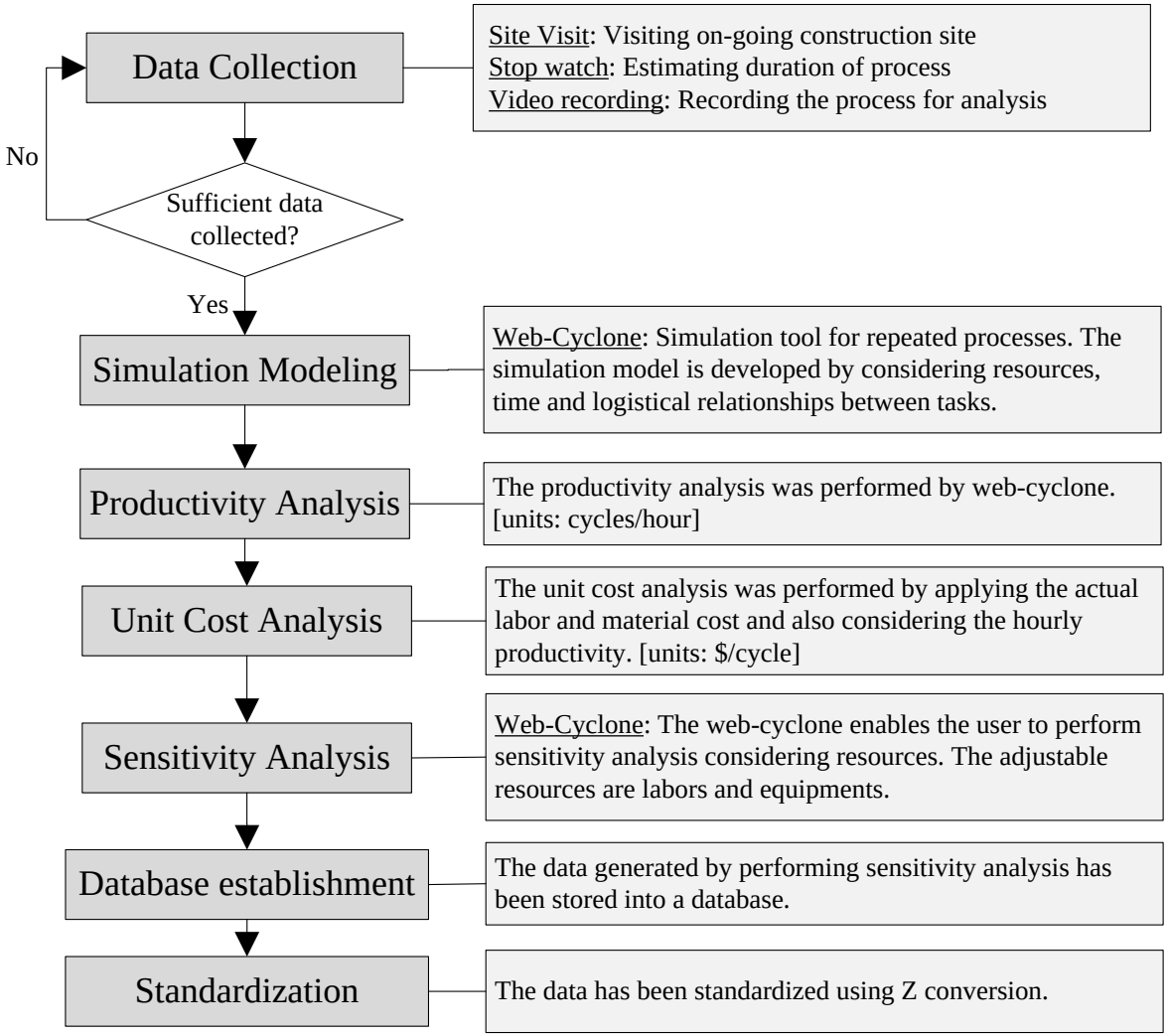
Correlation analysis between the word data and PMIS actual data has been performed.

Total 879 correlations which the correlation coefficient is higher than 0.7 has been found.

The meaning of each correlation should be analyzed for future work.

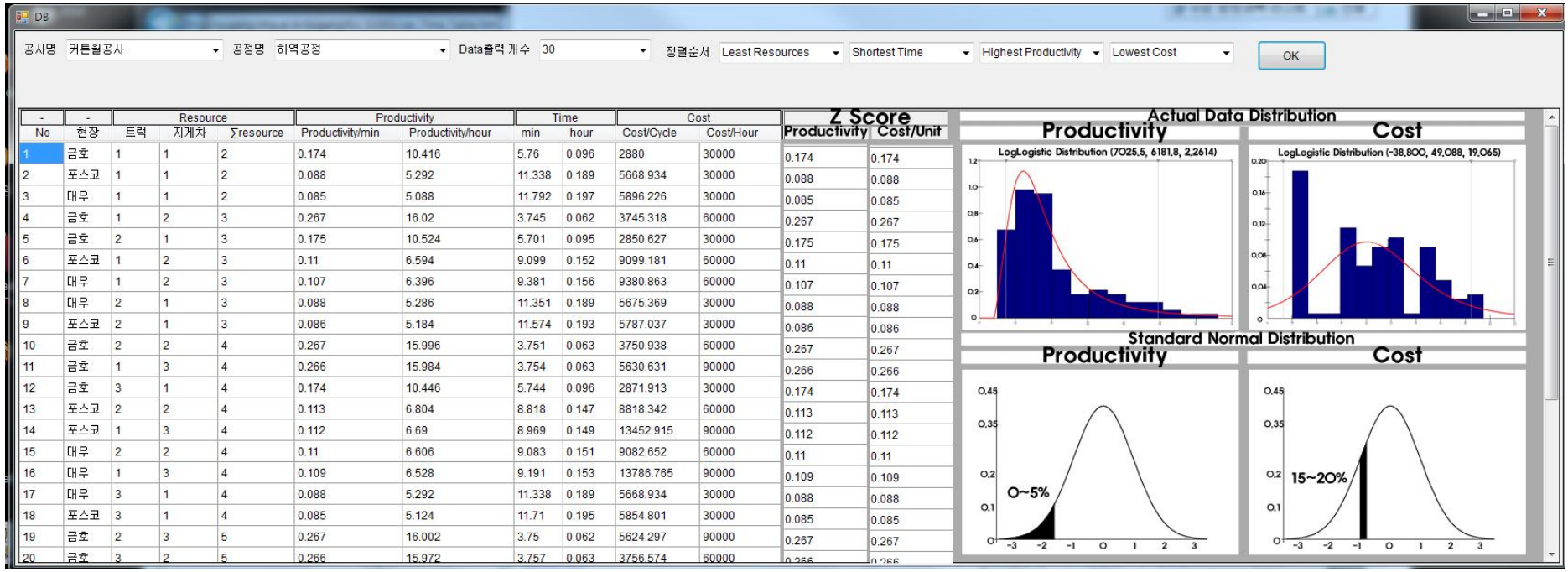
77	97	sum	미장26	0.704338	0
77	100	sum	sum	0.821362	0
77	122	sum	반장37	0.867587	0
77	139	sum	미장49	0.918647	0
77	155	sum	마감	0.720157	CHECK
77	156	sum	바닥	0.905607	CHECK
77	157	sum	배근작업	0.920721	CHECK
77	167	sum	미장sum	0.716617	CHECK
77	168	sum	방수	0.723242	CHECK
77	169	sum	방수sum	0.791864	CHECK
77	170	sum	작업	0.786195	CHECK
78	9	반장21	25-210-125	0.725001	0
78	69	반장21	유리	0.737675	0
78	70	반장21	기계설비	0.821058	0
78	71	반장21	닥트	0.894854	0
78	72	반장21	반장	0.872806	0
78	73	반장21	용접	0.855339	0
78	77	반장21	sum	0.843507	0
78	79	반장21	직원22	0.953356	0
78	80	반장21	금속	0.971942	0
78	81	반장21	sum	0.982097	0
78	90	반장21	전기설비	0.789253	0
78	94	반장21	sum	0.754725	0
78	97	반장21	미장26	0.828474	0
78	100	반장21	sum	0.797232	0
78	112	반장21	잡철물	0.762225	0
78	122	반장21	반장37	0.92403	0
78	123	반장21	수장	0.794372	0
78	125	반장21	sum	0.790979	0
78	139	반장21	미장49	0.826488	0
78	154	반장21	방수	0.885654	0
78	156	반장21	바닥	0.826675	CHECK
78	157	반장21	배근작업	0.839301	CHECK
78	168	반장21	방수	0.701581	CHECK
78	169	반장21	방수sum	0.764266	CHECK
78	170	반장21	작업	0.75813	CHECK
79	2	직원22	25-180-12	0.710533	0

DATA ASSESSMENT USING STATISTICAL APPROACHES



DATA ASSESSMENT USING STATISTICAL APPROACHES

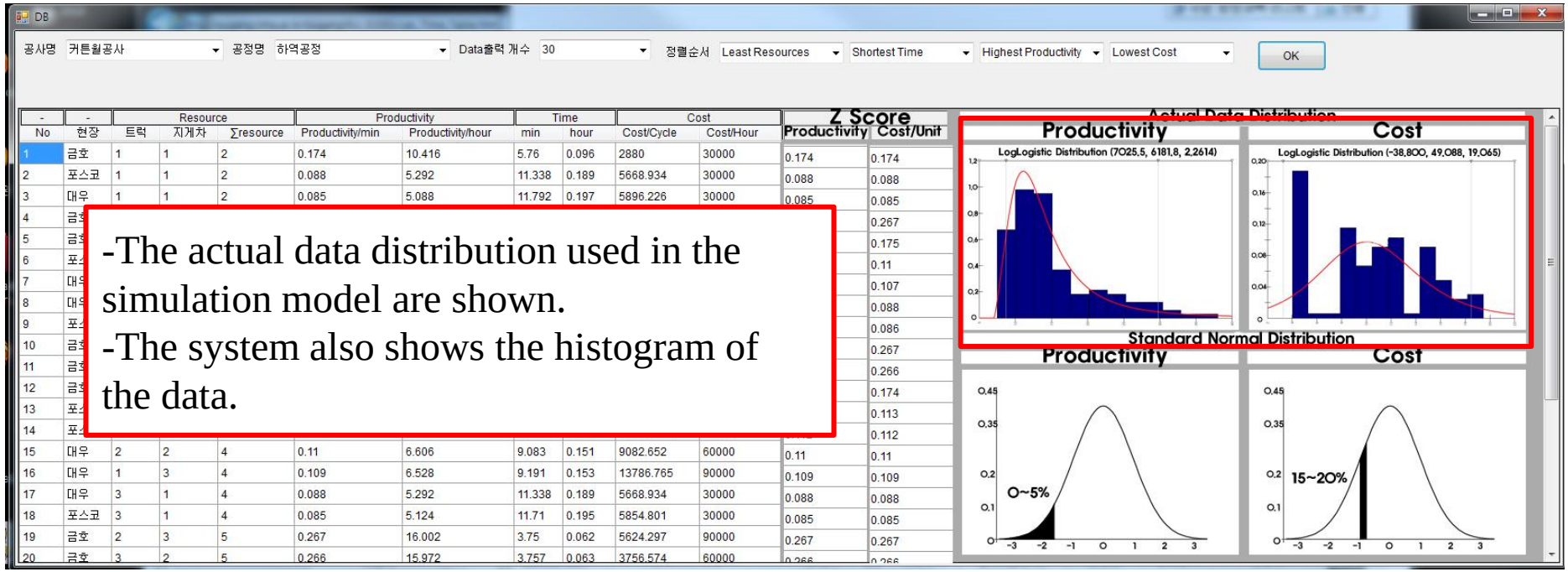
Overviews of the DB



-	-	Resource			Productivity		Time		Cost		Z Score	
No	현장	트럭	지게차	∑resource	Productivity/min	Productivity/hour	min	hour	Cost/Cycle	Cost/Hour	Productivity	Cost/Unit
1	금호	1	1	2	0.174	10.416	5.76	0.096	2880	30000	0.174	0.174
2	포스코	1	1	2	0.088	5.292	11.338	0.189	5668.934	30000	0.088	0.088
3	대우	1	1	2	0.085	5.088	11.792	0.197	5896.226	30000	0.085	0.085
4	금호	1	2	3	0.267	16.02	3.745	0.062	3745.318	60000	0.267	0.267
5	금호	2	1	3	0.175	10.524	5.701	0.095	2850.627	30000	0.175	0.175
6	포스코	1	2	3	0.11	6.594	9.099	0.152	9099.181	60000	0.11	0.11
7	대우	1	2	3	0.107	6.396	9.381	0.156	9380.863	60000	0.107	0.107
8	대우	2	1	3	0.088	5.286	11.351	0.189	5675.369	30000	0.088	0.088
9	포스코	2	1	3	0.086	5.184	11.574	0.193	5787.037	30000	0.086	0.086
10	금호	2	2	4	0.267	15.996	3.751	0.063	3750.938	60000	0.267	0.267
11	금호	1	3	4	0.266	15.984	3.754	0.063	5630.631	90000	0.266	0.266
12	금호	3	1	4	0.174	10.446	5.744	0.096	2871.913	30000	0.174	0.174
13	포스코	2	2	4	0.113	6.804	8.818	0.147	8818.342	60000	0.113	0.113
14	포스코	1	3	4	0.112	6.69	8.969	0.149	13452.915	90000	0.112	0.112
15	대우	2	2	4	0.11	6.606	9.083	0.151	9082.652	60000	0.11	0.11
16	대우	1	3	4	0.109	6.528	9.191	0.153	13786.765	90000	0.109	0.109
17	대우	3	1	4	0.088	5.292	11.338	0.189	5668.934	30000	0.088	0.088
18	포스코	3	1	4	0.085	5.124	11.71	0.195	5854.801	30000	0.085	0.085
19	금호	2	3	5	0.267	16.002	3.75	0.062	5624.297	90000	0.267	0.267
20	금호	3	2	5	0.266	15.972	3.757	0.063	3756.574	60000	0.266	

DATA ASSESSMENT USING STATISTICAL APPROACHES

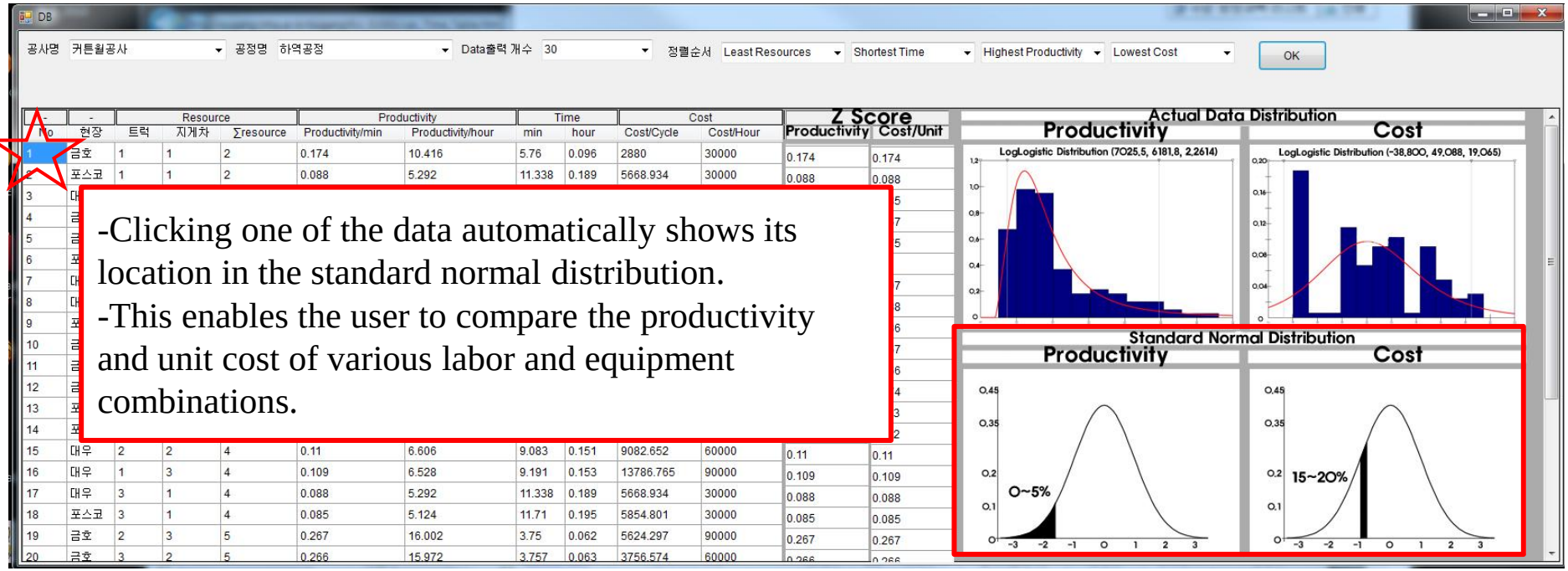
Illustration of Data Distribution



No	현장	Resource 트럭	지게차	∑resource	Productivity Productivity/min	Productivity/hour	Time min	hour	Cost Cost/Cycle	Cost/Hour	Z Score Productivity	Cost/Unit
1	금호	1	1	2	0.174	10.416	5.76	0.096	2880	30000	0.174	0.174
2	포스코	1	1	2	0.088	5.292	11.338	0.189	5668.934	30000	0.088	0.088
3	대우	1	1	2	0.085	5.088	11.792	0.197	5896.226	30000	0.085	0.085
4	금호											0.267
5	금호											0.175
6	포스											0.11
7	대우											0.107
8	대우											0.088
9	포스											0.086
10	금호											0.267
11	금호											0.266
12	금호											0.174
13	포스											0.113
14	포스											0.112
15	대우	2	2	4	0.11	6.606	9.083	0.151	9082.652	60000	0.11	0.11
16	대우	1	3	4	0.109	6.528	9.191	0.153	13786.765	90000	0.109	0.109
17	대우	3	1	4	0.088	5.292	11.338	0.189	5668.934	30000	0.088	0.088
18	포스코	3	1	4	0.085	5.124	11.71	0.195	5854.801	30000	0.085	0.085
19	금호	2	3	5	0.267	16.002	3.75	0.062	5624.297	90000	0.267	0.267
20	금호	3	2	5	0.266	15.972	3.757	0.063	3756.574	60000	0.266	

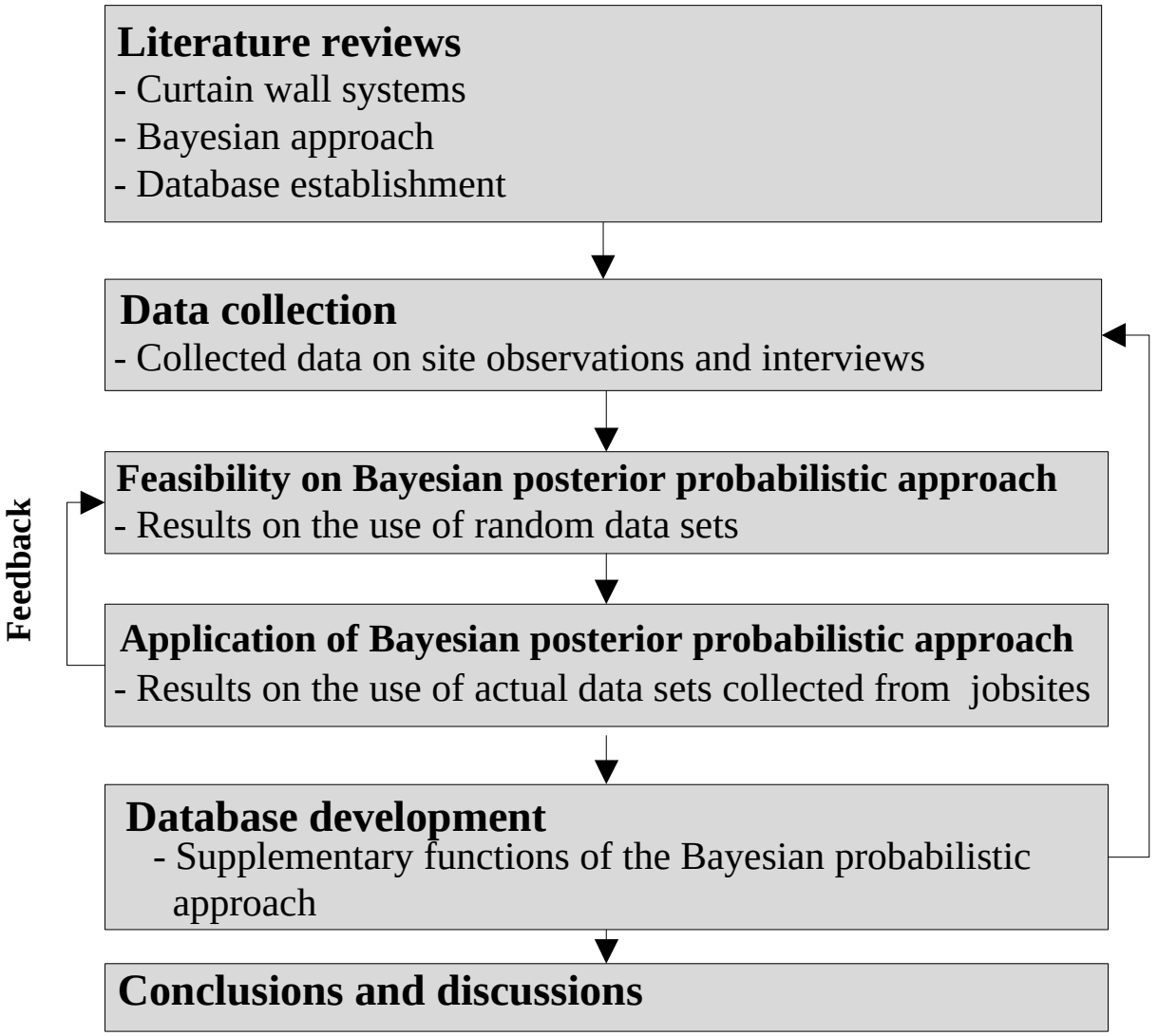
DATA ASSESSMENT USING STATISTICAL APPROACHES

Illustration of Statistical Position



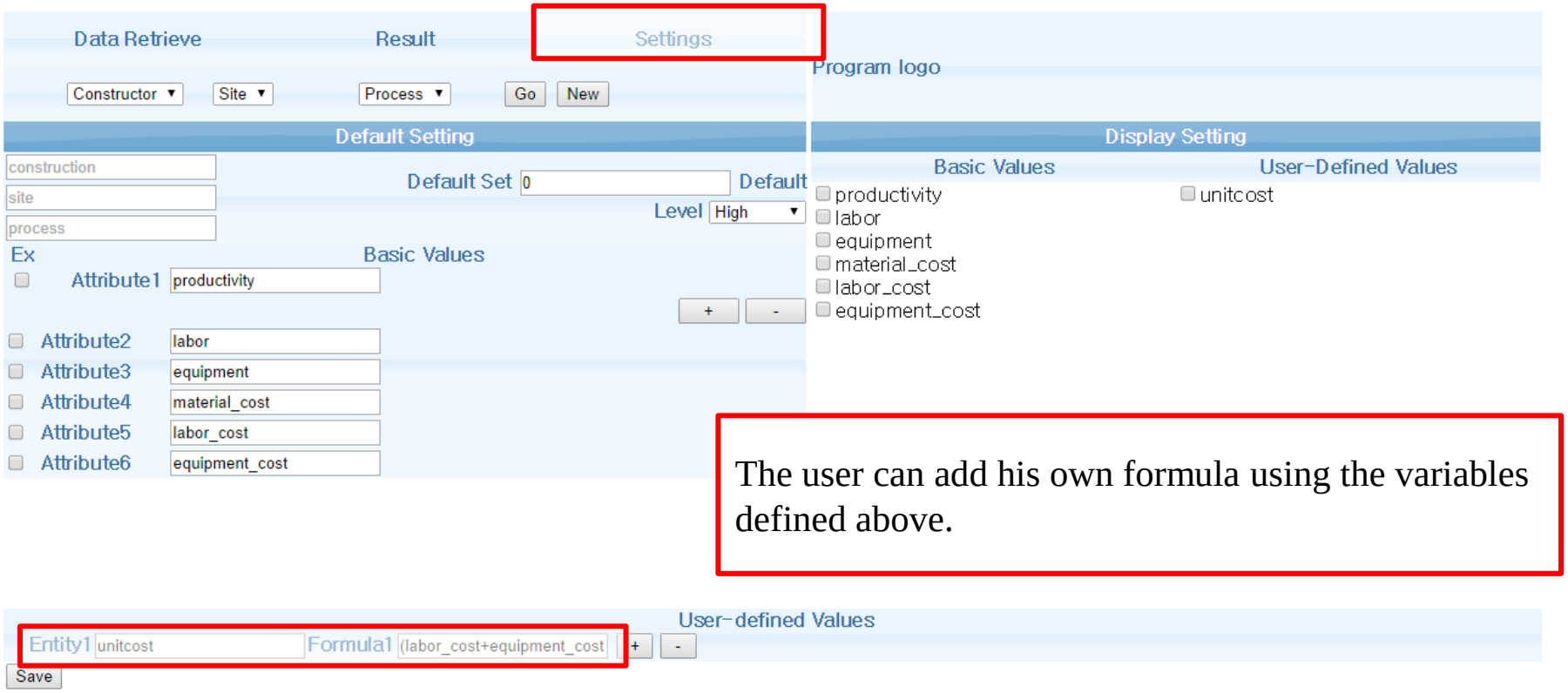
INFORMATION PREDICTION USING STATISTICAL APPROACHES

Research method



INFORMATION PREDICTION USING STATISTICAL APPROACHES

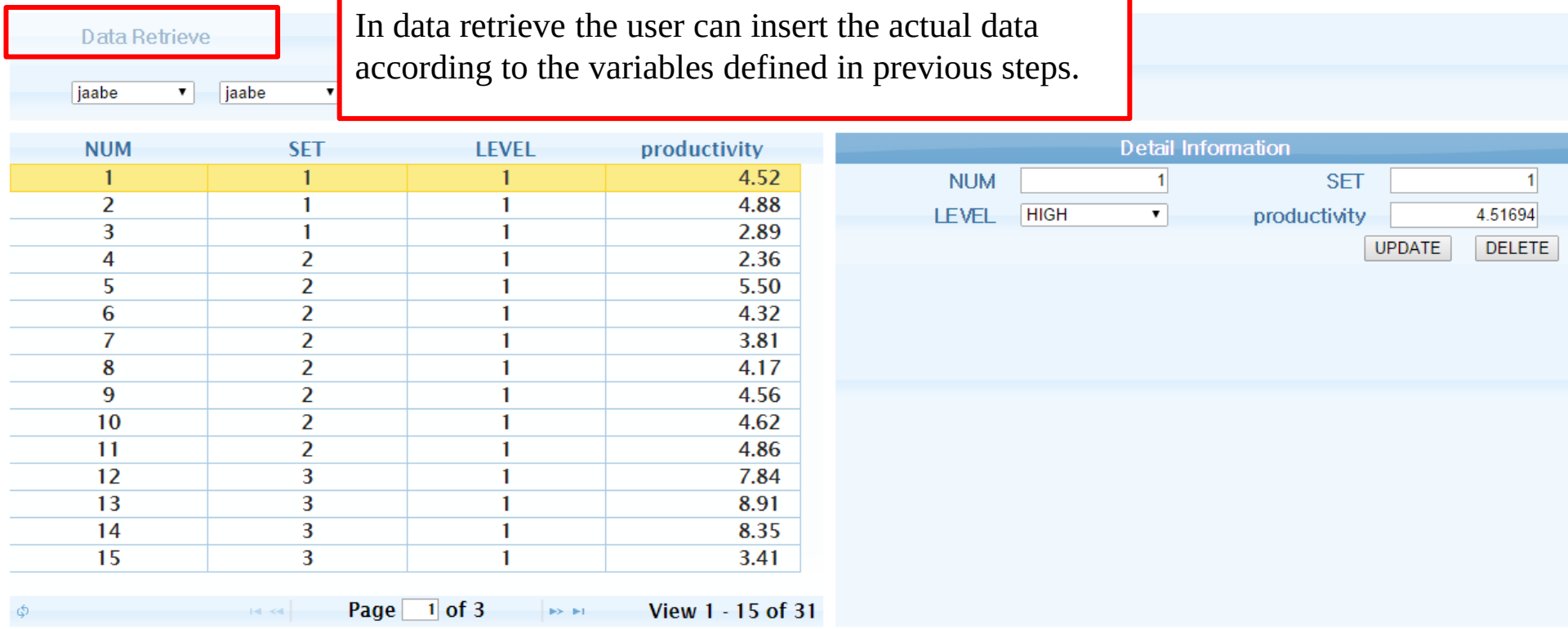
Interface for Data Input



INFORMATION PREDICTION USING STATISTICAL APPROACHES

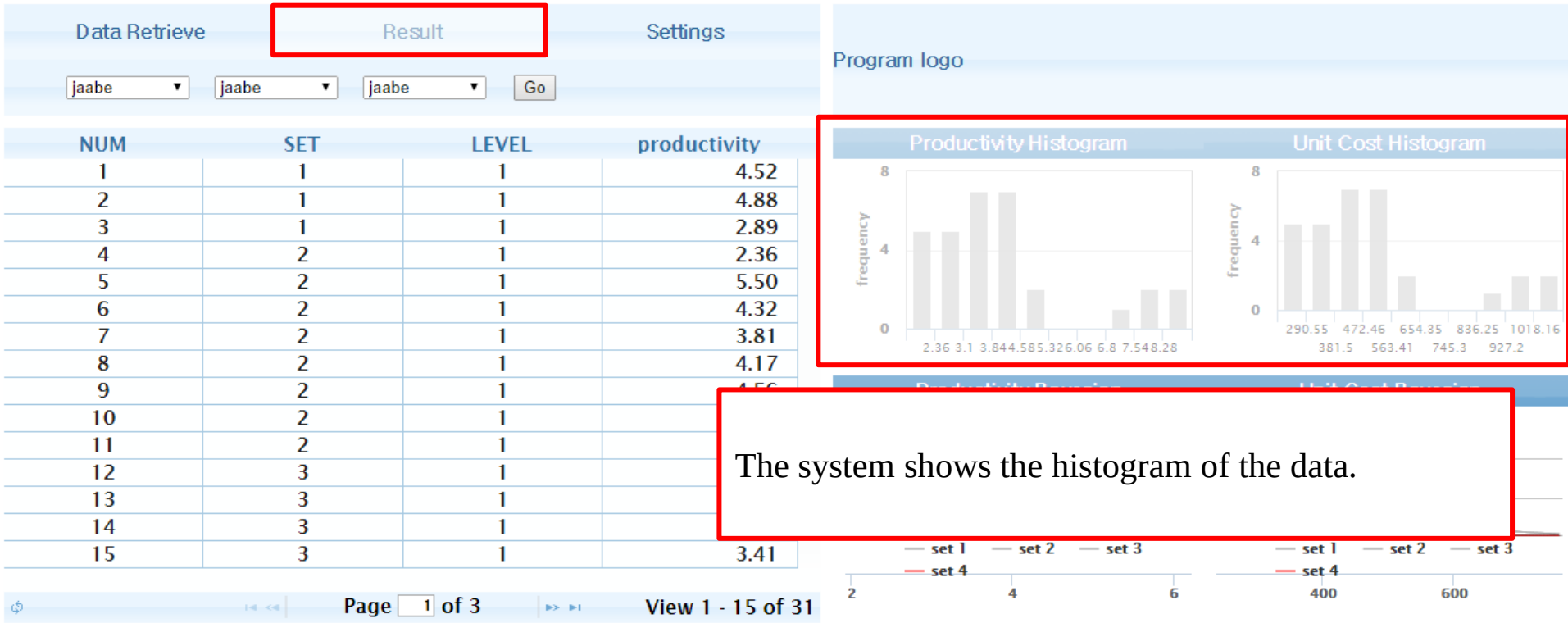
Interface for Data Retrieve

In data retrieve the user can insert the actual data according to the variables defined in previous steps.



INFORMATION PREDICTION USING STATISTICAL APPROACHES

Illustration of the Analyzed Results



The system shows the histogram of the data.

INFORMATION PREDICTION USING STATISTICAL APPROACHES

Illustration of the Analyzed Results

Data Retrieve | Result | Settings

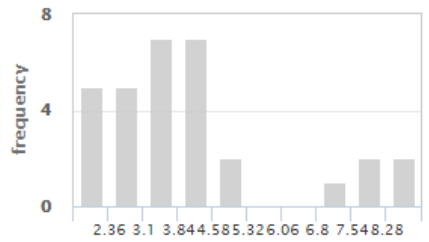
jaabe ▾ jaabe ▾ jaabe ▾ Go

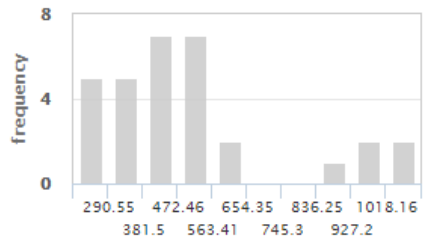
Program logo

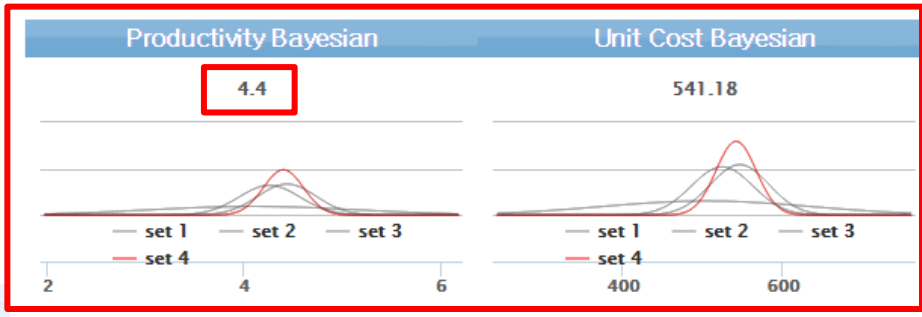
NUM	SET	LEVEL	productivity
1	1	1	4.52
2	1	1	4.88
13	3	1	8.91
14	3	1	8.35
15	3	1	3.41

Page 1 of 3 View 1 - 15 of 31

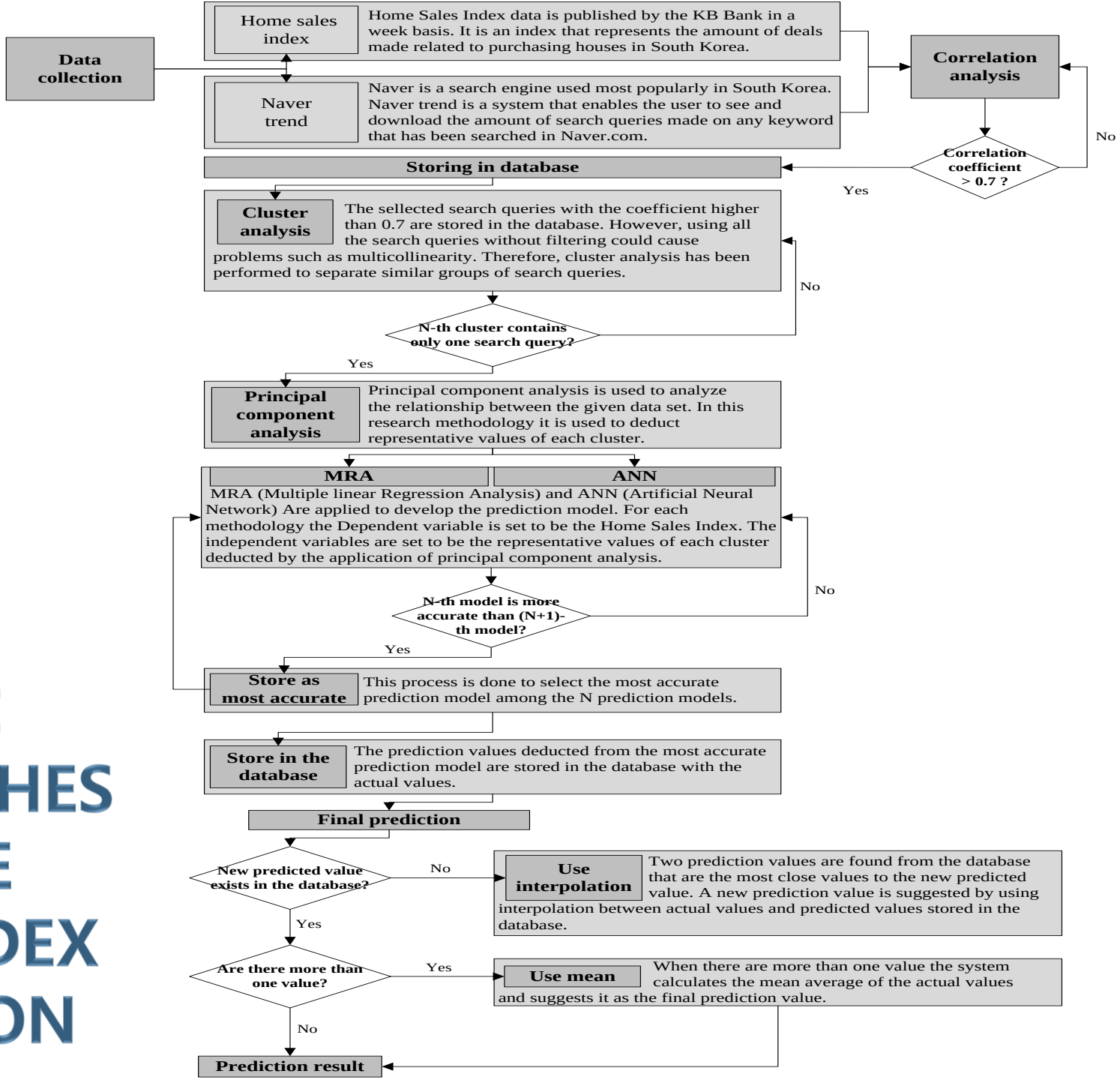
This is the analysis results using the Bayesian probabilistic approach.
In this example there are four data sets, therefore, it can be seen that there are four distributions.
The red distribution represents the result.





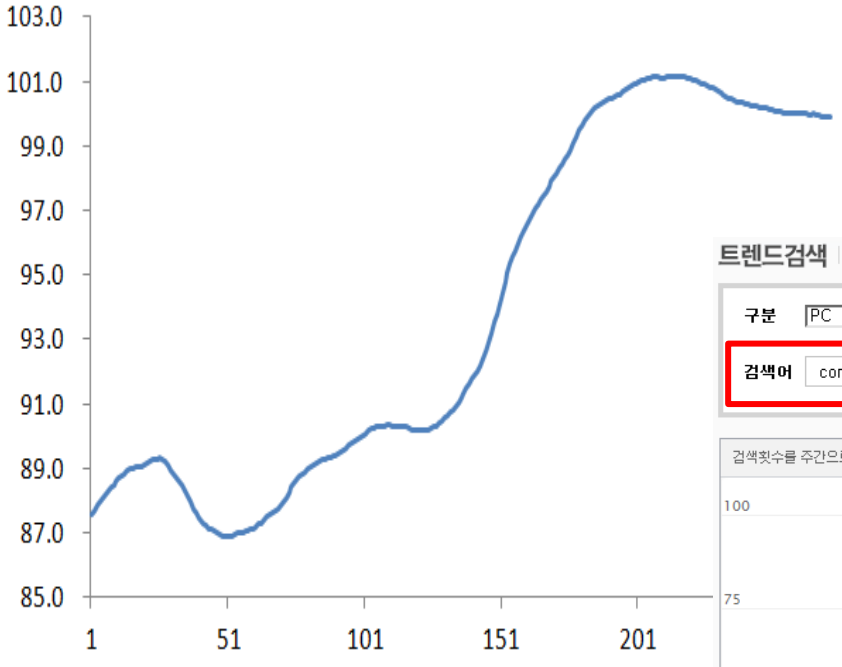


BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION



BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

HSI Data



Naver Trend

BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

Selected Search Queries

Search Queries				
2금융권	신용조회	오피스텔	국민주택채권	등기부등본
중개수수료	중고차	닥터아파트	은행대출	전세계약서
주공아파트	IBK기업은행	민원24	소형주택	주민등록증진위확인
광주집값	대구집값	동탄신도시	분양권프리미엄	신혼부부 집
계약면적	등기사항전부증명서	부동산조회	매도용인감증명서	보증금반환
사다리차비용	소득증빙서류	수익형부동산		

Correlation Analysis

search queries	상관계수	유의확률
1금융권	.192	.002
2금융권	-.765	.000
3금융권	.528	.000
IBK기업은행	.858	.000
계약면적	.837	.000
광주집값	.706	.000
국민주택채권	-.831	.000
닥터아파트	-.905	.000
대구집값	.778	.000
대전집값	.284	.000
동탄신도시	-.816	.000
등기부등본	-.825	.000
등기사항전부증명서	.849	.000
매도용인감증명서	.835	.000
민원24	.858	.000
보증금반환	.724	.000
부동산조회	.858	.000
부산집값	.571	.000
분양권	-.576	.000
분양권매매	.441	.000
분양권프리미엄	.719	.000
사다리차비용	.833	.000
서울집값	-.365	.000
소득증빙서류	.893	.000
소형주택	.775	.000
수익형부동산	.888	.000
수익형부동산	.888	.000
신용조회	-.898	.000
신혼부부아파트	-.245	.000
신혼부부전세	.413	.000

BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

Cluster Analysis

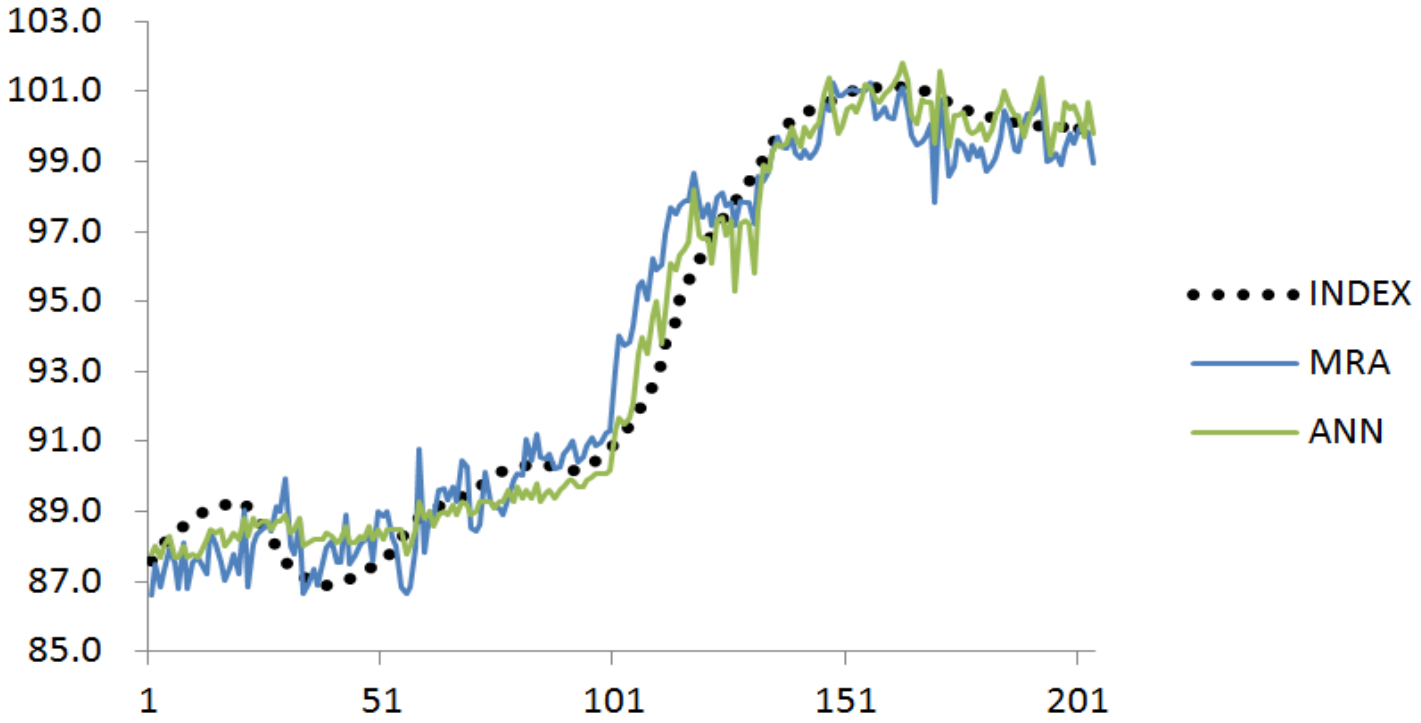
search queries	7	6	5	4	3	2
2금융권	1	1	1	1	1	1
신용조회	1	1	1	1	1	1
오피스텔	1	1	1	1	1	1
국민주택채권	5	5	4	4	1	1
등기부등본	5	5	4	4	1	1
중개수수료	5	5	4	4	1	1
중고차	5	5	4	4	1	1
닥터아파트	5	5	4	4	1	1
은행대출	5	5	4	4	1	1
전세계약서	5	5	4	4	1	1
주공아파트	5	5	4	4	1	1
IBK기업은행	2	2	2	2	2	2
민원24	2	2	2	2	2	2
소형주택	2	2	2	2	2	2
주민등록증진위확인	2	2	2	2	2	2
계약면적	**3**	**3**	**3**	3	3	2
광주집값	4	4	2	2	2	2
대구집값	4	4	2	2	2	2
동탄신도시	4	4	2	2	2	2

Principal Component Analysis

Num	index	1 cluster	2 cluster	4 cluster	5 cluster
1	87.6	1.04977	-1.18136	1.58591	-1.20381
2	88.1	0.717	-1.14796	1.32595	-1.26092
3	88.5	1.08544	-1.26073	1.67132	-0.92794
4	88.8	0.43747	-1.291	1.27459	-1.20788
5	89.0	1.21434	-1.24213	1.04835	-1.3545
6	89.2	1.67226	-1.05924	1.07667	-1.07133
7	89.3	1.3468	-1.16443	0.91849	-1.18628
8	89.1	1.65456	-1.07098	1.1521	-1.38944
9	88.6	1.33961	-1.12444	0.59324	-1.17678
10	88.0	0.47392	-0.98518	0.68877	-1.13361
11	87.4	1.1357	-1.07778	1.03508	-1.054
			.....		
200	100.0	-1.11092	1.11627	-1.0566	1.45215
201	100.0	-1.23976	1.02335	-0.96124	0.68083
202	100.0	-1.3686	0.69108	-1.37334	0.56646
203	99.9	-1.16822	0.82175	-1.02179	1.32825

BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

Multiple Linear Regression & Artificial Neural Network



BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

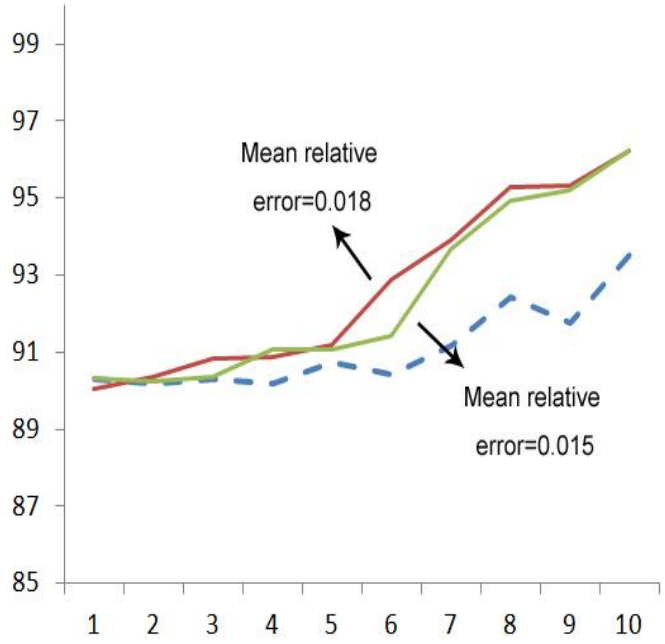
Estimation using Interpolation

Actual data	MRA				ANN			
	Normal	Error	Interpolation	Error	Normal	Error	Interpolation	Error
90.32	90.04	0.0030	90.31	0.0001	89.40	0.0101	90.30	0.0002
90.19	**90.37**	**0.0020**	**90.26**	0.0008	89.40	0.0088	90.30	0.0012
90.30	90.83	0.0058	90.38	0.0010	90.30	0.0000	90.21	0.0010
90.18	90.88	0.0077	91.07	0.0098	89.40	0.0087	90.30	0.0013
90.74	91.19	0.0049	91.06	0.0035	90.30	0.0048	90.21	0.0058
90.40	92.88	0.0274	91.41	0.0112	90.40	0.0000	90.22	0.0020
91.18	93.92	0.0301	93.66	0.0272	90.40	0.0086	90.22	0.0105
92.42	95.28	0.0309	94.91	0.0269	92.40	0.0002	92.23	0.0021
91.75	95.33	0.0390	95.21	0.0376	100.20	0.0921	100.20	0.0921
93.52	96.24	0.0291	96.23	0.0290	93.50	0.0002	93.50	0.0002

Calculation: $\frac{(90.33-90.25)}{(90.46-90.21)} \times (90.37-90.21)+90.21=90.26$

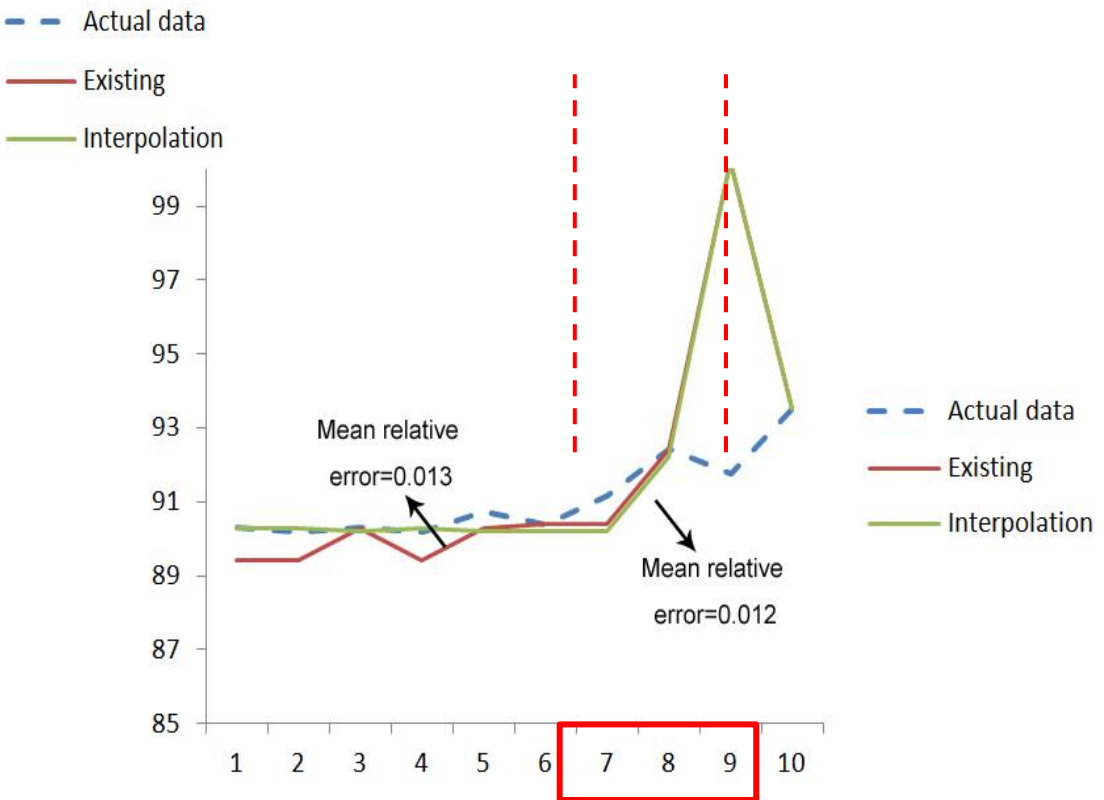
BIG DATA ANALYTIC APPROACHES ON HOME SALES INDEX PREDICTION

Result Comparison



Multiple Linear Regression

Artificial Neural Network



인하대학교
INHA UNIVERSITY

빅데이터 전문인력 양성 협약: 인하대학교 – 한국DB진흥원
소프트웨어융합공학 연계전공: 인하대학교 - 삼성전자