

## 04 질병예측



# 질병예측 필요성과 활용가능 데이터



# 1 질병예측 필요성

- 질병으로 인한 개인/사회적 비용증가
- 고연령화로 인해 질병으로 인한 기회비용 대응필요
- 건강 및 의료관련 데이터 체계화에 따른 높은 가치창출 기회
- 신산업 육성 및 활용 시 세계적 경쟁력 가능성 높음



## 2 예측을 위한 활용가능 데이터

### ✓ 건강검진 데이터

- 전국민 대상 1, 2년 주기 데이터 가용

### ✓ 진료 데이터

- 전국민 데이터 가용

### ✓ 검사 데이터

- 다양한 검사가 실시되어 활용 시 가치 높음



## 2 예측을 위한 활용가능 데이터

### ✓ 투약 데이터

- 전국민 데이터 가용

### ✓ 건강관리 IoT 및 App 데이터

- 급증하고 있는 데이터
- 아직 누적분량 및 기간 부족하나 향후 활용성 높음



### 3 예측을 위한 활용가능 데이터

[건강보험 청구자료 150 TB]

명세서 일반내역	20TB	환자 및 요양기관 정보, 내원일수, 주상병
진료내역	30TB	처치, 수술, 병원내약처방, 투여량 등 상세의료 서비스 자료
수신자 상병내역	40TB	주상병, 부상병 및 추가 부상병
처방전교부 상세내역	60TB	원외처방 약에 대한 정보



## 2 예측을 위한 활용가능 데이터

- 환자, 질병, 병원 매출 추이
- 질병 별 평균 진료기간, 비용 분석
- 전염성 질환 확산 분석
- 환자특성 별 투약효과 분석
- 불법 처방분석(Fraud Detection)



## 2 예측을 위한 활용가능 데이터

[코호트 DB(2002~2009)]

자격 및 보험료 자료	건강보험 가입자 성/연령, 거주지, 소득 분위
진료상세 자료	건강보험 청구자료
건강검진 자료	최소 5회 이상 건강검진 받은 수검자 (해당인원 매년 약 6만명 수준발생) 검진결과와 문진내용 (가족력, 흡연, 음주, 운동 등)



## 2 예측을 위한 활용가능 데이터

- 환자중심의 데이터로 **100만 고객층화 추출**되어 활용성 높음
- 건강보험 청구자료 기준 외의 소득, 생활습관, 가족력 등 **다양한 추가정보 가용**
- 질병, 생활습관과 연계 → **질병예측에 유의미한 변수** 제공
- 질병예측에 **효율적인 정보 소스**  
예) 당뇨예측, 암 발생 예측



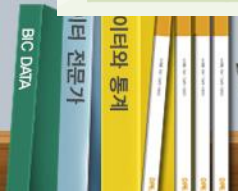


## 2 예측을 위한 활용가능 데이터

### 건강검진자료(문진포함)-1

h-well  
국민건강보험

연번	변수명	변수형태	연번	변수명	변수형태
1	검진년도	char(4)	11	총콜레스테롤	char(4)
2	검진기관대체번호	char(8)	12	트리글리세라이드	char(4)
3	검진기관종별코드	char(4)	13	<b>HDL</b> 콜레스테롤	char(4)
4	검진기관 시도코드	char(2)	14	<b>LDL</b> 콜레스테롤	char(4)
5	신장	char(3)	15	혈색소	char(3)
6	체중	char(3)	16	혈청크레아티닌	char(3)
7	허리둘레	char(3)	17	(혈청지오티) <b>ASL</b>	char(4)
8	혈압최고	char(3)	18	(혈청지피티) <b>ALT</b>	char(4)
9	혈압최저	char(3)	19	감마지피티	char(3)
10	식전혈당	char(3)	20	검진일자	char(8)



# 2 예측을 위한 활용가능 데이터

## 건강검진자료(문진포함)-2

h-well  
국민건강보험

연번	변수명		상세변수		
22	과거병력	진단여부	1.뇌졸중(중풍)	4.당뇨병	6.폐결핵
			2.심장병(심근경색/협심증)	5.이상지혈증	7.기타(암포함)
			3.고혈압		
23		약물치료여부	1.뇌졸중(중풍)	4.당뇨병	6.폐결핵
			2.심장병(심근경색/협심증)	5.이상지혈증	7.기타(암포함)
			3.고혈압		
24	가족력		1.뇌졸중(중풍)	3.고혈압	5.기타(암포함)
			2.심장병(심근경색/협심증)	4.당뇨병	
25	흡연	4-1)지금까지 평생 총 5갑(100개비)이상의 담배를 피운 적이 있습니까?			
		4-2)과거 흡연을 었으나 현재는 끊으셨다면	몇 년째	평균 하루 흡연량	
		4-3)현재도 흡연을 하신다면	몇 년째	하루 흡연량	
26	음주(술) 습관	5-1)1주에 평균 며칠이나 술을 마십니까			
		5-2)술을 드실때 보통 하루에 얼마나 마십니까 (술 종류에 관계없이)			
27	신체활동(운동)	6-1)20분이상 일주일에 몇회정도(고강도)			
		6-2)30분이상 일주일에 몇회정도(중등도)			
		6-3)30분이상 일주일에 몇회정도(걷기)			



## 04 질병예측

2

# 모델링 접근 및 활용방안



# 1 모델링 접근방안

당뇨이력 있는 사람

X

당뇨이력이 없는 사람

X

당뇨발생

당뇨발생  
안함

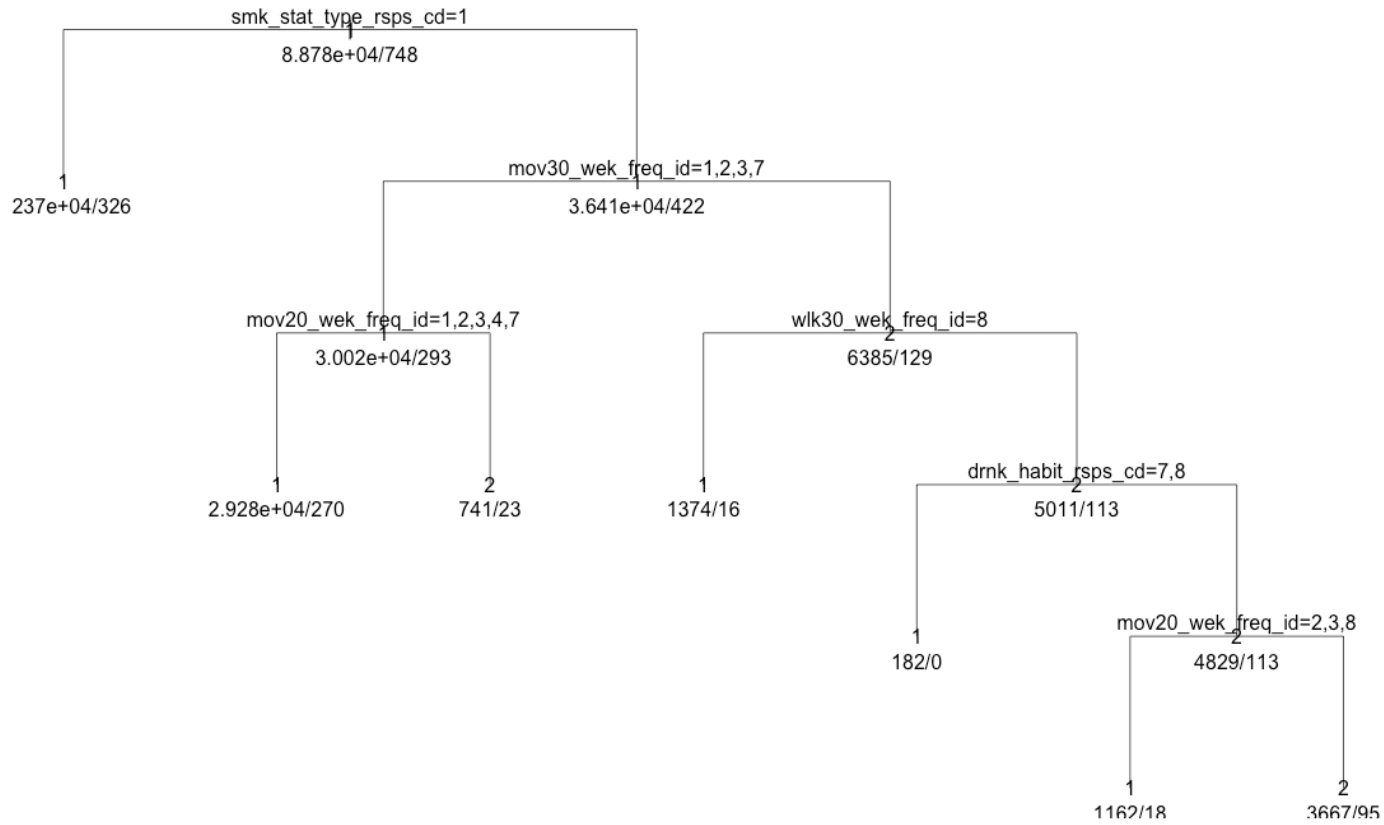
과거 7년간의 데이터

1년간의  
데이터



## 2 Prototype Model 예시(Classification)

Classification Tree for diabet



### 3 활용방안

- 질병 발생 1년 전, 향후 1년 내 질병 걸릴 확률 제공 : 맞춤형 건강관리/검진계획 제공 → 환자만족도 향상
- 질병 예측에 따른 처방 : **대환자 서비스 개선**
- 질병 예측에 따른 건강검진 계획 : **의료보험 비용절감**
- 질병에 대한 예측 및 인과관계 규명을 통한 **의료발전**

