<1페이지>

발표 시작하겠습니다.

저희 제목은 ‘디지털 병간호 서비스’이며 조장 및 발표를 맡게 된 15학번 이민혁입니다.

<2페이지>

목차는 앞에 보이시는 것과 같습니다.

<3페이지> 아이디어 선정 배경

저희 아이디어는 첫번째로 이번 코로나 사태와 연관이 있습니다. 중국에서 발생한 코로나19 바이러스로 인해 오른쪽 뉴스 기사를 보시 듯이 국내 많은 병원들이 병원 내 감염 예방을 위해 병문안을 통제하였습니다. 두번째로 오랜 기간동안 입원해야 하는 환자의 가족들은 여러 사정으로 환자의 곁을 계속 지키기에는 부담이 생기는 경우가 많다는 얘기를 들었습니다. 저희 조는 이 두가지 배경을 바탕으로 ‘디지털 병간호 서비스’라는 아이디어를 생각했습니다.

<4페이지> 아이디어 목적

저희 아이디어의 목적은 첫번째, 보호자가 병원 방문을 하지 않더라도 입원한 환자의 건강 상태를 확인할 수 있어야 하고, 두번째, 환자의 누적 건강 상태 데이터를 분석하여 앞으로의 건강 상태와 다른 질병 발생 가능성을 예측할 수 있어야 합니다. 세번째로 병원에서 제공하는 환자의 식단을 분석하여 섭취한 영양 성분 정보를 제공할 수 있어야 하고 마지막으로 환자가 복용한 약품명과 성분의 정보를 제공받을 수 있어야 합니다.

<5페이지> 데이터 수집

앞서 아이디어 목적을 바탕으로 필요한 데이터를 수집하였습니다. 공공데이터 포털을 이용하여 병원에서 진료한 과거 환자의 정보, 음식 영양소 정보, 약 성분 정보들을 가져왔고, 실제 병원 식단을 참고하여 영양 균형이 잡힌 식단 표 데이터와 가상의 현재 환자 데이터를 생성하였습니다.

<6페이지> 데이터 수집

앞서 설명들은 데이터들입니다.

<7,8페이지> 데이터 모델링

데이터 베이스 설계의 첫번째, 요구 사항 명세서입니다.

앞서 말한 아이디어 목적 4가지를 생각하고, 데이터베이스의 용도를 생각하며 요구 사항 명세서를 작성하였습니다.

<9페이지> 데이터 모델링

데이터 베이스 설계의 두번째, 개념적 설계 작업 과정입니다.

지금 보시는 표는 요구 사항 분석 결과를 기반으로 중요한 개체를 추출하고 각 개체의 주요 속성과 키 속성을 선별한 것입니다. 진한 파란색 안에 들어있는 것이 기본키이며, 옅은 파란색 안에 들어있는 것들이 그 개체의 주요 속성들입니다.

<10페이지>

그 다음 개체 간의 관계를 생각하며 관계를 추출하였습니다.

현재환자와 현재환자측정의 관계에 대해 설명해 드리겠습니다.

현재환자는 측정 테이블을 여러 개 가질 수 있으며 optional한 관계입니다.

현재환자 측정 값마다 한 명의 환자와 연결되어야 합니다.

<11페이지>

현재환자는 여러가지 약을 복용할 수 있으며, Optional 함으로 복용하지 않을 수도 있고,

약은 여러 명의 환자에게 복용되어 질 수 있습니다.

<12페이지>

현재환자는 여러가지 식단을 먹을 수 있으며, 먹지 않을 수도 있습니다.

한가지 식단은 여러 환자에게 제공되어 질 수 있습니다.

식단영양소에 있는 음식은 식단표에 여러개 들어가 있으며 들어가지 않아도 됩니다.

식단표에 있는 음식은 반드시 식단 영양소에 있는 음식이어야만 합니다.

<13페이지>

앞에서 말한 관계들을 ER 다이어그램으로

DB를 구축한 결과 다음과 같은 테이블이 나오게 되었습니다.

지금부터 DB구축한 과정에 대해서 설명드리겠습니다.

<14페이지>

현재환자 테이블은 기본키가 환자ID이고 나이가 0보다 커야 합니다.

<15페이지>

현재환자 측정 테이블은 기본키가 환자ID와 검사 날짜의 복합키이고, 현재환자 테이블의 환자ID속성을 참조합니다.

<16페이지>

과거환자 테이블은 기본키가 과거환자ID입니다.

<17페이지>

식사 테이블은 기본키가 환자ID, 검사날짜, 식사시간의 복합키이고, 식단표의 식단ID와 현재환자의 환자ID를 참조합니다.

<18페이지>

식단표 테이블은 기본키가 식단 아이디와 음식명의 복합키이고, 식단 영양소의 음식명을 참조합니다.

<19페이지>

식단영양소 테이블은 기본키가 음식명입니다.

<20페이지>

약이름 테이블은 기본키가 질병명이고 약성분의 약이름과 현재환자의 질병명 참조합니다.

<21페이지>

약성분 테이블은 기본키가 약 이름입니다.

<22페이지>

그 다음으로 DML를 이용하여 데이터를 추출해 보았습니다.

앞에 보이시는 표는 ‘석민규’환자의 2020년 6월8일 드셨던 음식의 영양 성분들입니다.

<23페이지>

그 다음으로 ‘석민규’환자의 측정 데이터입니다. 이 측정데이터들은 후에 이 환자가 다른 질병에 걸릴지 안 걸릴지를 예측하는 데이터입니다.

<24페이지>

다음으로는 ‘석민규’환자가 처방받은 약에 대한 내용입니다.

<25페이지>

환자의 검진 정보를 이용하여 앞으로의 질병 발생 가능성을 예측할 수 있도록, 과거 환자 데이터를 이용하여 로지스틱 회귀분석을 하였습니다. 먼저 32개의 환자 검진 정보 속성과 26개의 질병 발생 여부 데이터가 있는 과거 환자 데이터를 불러온 후, 질병 발생 여부를 종속 변수로, 환자 검진 정보를 독립 변수로 설정합니다. 그 후 학습 데이터 80%, 테스트 데이터 20%로 설정한 후 26개의 질병 각각에 대하여 로지스틱 회귀 분석을 실행합니다. 이 때 각각에 모델에 대한 상관계수를 따로 테이블로 만들어서 저장합니다.

<26페이지>

만들어진 상관 계수 테이블은 각각에 질병 발생 여부에 대한 상관계수가 저장되어 있습니다. 이 데이터에서 각 질병별로 상관계수 상위 30%만 색깔을 달리하였습니다.

<27페이지>

그 후 가장 많이 중복된 변수 10가지를 선정하였습니다. 이렇게 선정된 변수들은 아래와 같고 이독립변수를 32개에서 10개로 줄인 이유는, 환자 개인이 자신의 검사 정보를 가지고 다른 질병 발생가능성을 알고 싶어할 때 32개의 변수 모두 입력하는 것은 어렵기 때문에 공통적으로 영향을 많이 주는 변수만을 선정하였습니다.

<28페이지>

이제 줄어든 10개의 환자 정보속성을 가지고 다시한번 26번의 로지스틱 회귀분석을 실행합니다. 각 모델의 정확도를 출력해보았는데 대부분의 질병 예측 정확도가 90%이상인 예측 모델이 생성되었습니다.

<29페이지>

각 모델의 신뢰도를 알아보기 위해 ROC커브 그려보고 AUC를 계산해보았습니다.

모든 모델에 대한 ROC커브를 PPT에 실을 수 없어 고혈압과 위암, 두 가지만 선정하였습니다.

AUC는 각각 0.816과 0.768이고 PPT에 첨부하지는 않았지만 모든 모델은 0.7이상의 AUC를 가지고 있어 이 예측모델은 신뢰할 수 있습니다.

<30페이지>

이제 현재 환자 테이블에서 예시로 ‘석민규’환자의 질병 발생 가능성을 예측해보았습니다. 그리고 각각의 질병 발생 확률을 테이블로 만들고 내림차순으로 정렬하였습니다.

<31페이지>

그 중 상위 5개를 선정하여 그래프로 나타내었는데, 대부분에 질병 발생 확률이 시간이 지남에 따라 감소하는 추세를 보이므로 ‘석민규’환자의 건강 상태는 호전되고 있다고 할 수 있습니다.

<32페이지>

마지막으로 저희가 만든 데이터 베이스를 C# Window form을 이용하여 구축해보았습니다.

환자의 아이디를 입력하면 그 환자의 ‘현재 건강 상태’,’건강 상태기록’, 질병 발생확률, 식단, 영양성분, 처방받은 약 성분등을 확인 할 수 있습니다.