

빅데이터를 활용한 AI 솔루션

HealthGuard : 건강 위험도 예측 시스템

LMS3차 유시온

Contents

01 프로젝트 개요

02 프로젝트 수행 절차 및 방법

03 프로젝트 수행 결과

04 자체 평가 의견

01 프로젝트 개요

- 주제

빅데이터, 딥 러닝을 활용하여 개인 건강 위험도 예측 프로그램 구현

- 기획 의도

빅데이터를 활용한, 딥 러닝 기법을 통해 검진 결과 입력 시 뇌·심혈관, 고혈압, 당뇨 위험도를 예측하는 프로그램 개발

- 개발 환경

Windows 10 Home / Visual Studio 2022 / C# / WPF / Python
Ubuntu 16.40 Linux OS / Visual Studio Code / Maria DB

02 프로젝트 수행 절차

구 분	기 간	활 동
사전 기획 및 데이터 수집	07/15 ~ 07/16	• 프로젝트 기획 및 주제 선정 및 개발 계획서 작성, 필요 데이터 수집 및 데이터 전처리
딥 러닝 학습	07/15 ~ 07/25	• 딥 러닝 기본적인 개념, 모델 학습
AI 모델 생성 및 학습	07/17 ~ 07/19	• 빅 데이터 활용하여 AI 모델 생성하여 학습
서버 구현	07/20 ~ 07/22	• 클라이언트 C#과 연결, mariaDB 연동 서버 구현, AI파트 Python과 연결
UI 및 클라이언트 구현	07/23 ~ 07/24	• WPF 사용하여 UI 구현, C++ 서버와 연결하는 클라이언트 구현
Python 서버 구현	07/23 ~ 07/24	• C++ 클라이언트 파트와 연결하는 서버 구현
테스트 및 오류 수정	07/25 ~ 07/26	• 테스트 및 디버깅 하여 오류 수정

03 프로젝트 수행 결과

✓ 데이터 :

혈압, 공복혈당, 총 콜레스테롤, 중성지방, 고밀도 콜레스테롤, 심장질환 여부, 뇌질환 여부, 혈압 여부, 당뇨 여부 등 각 검사 데이터 기록이 있는 백만개 데이터 사용

✓ 모델 주요 구성 요소 :

- Dense 레이어 : 128, 64, 32개의 뉴런을 가진 레이어 사용
- 활성화 함수: 각 Dense 레이어에서 ReLU함수 사용.
- 배치 정규화 (Batch Normalization): 각 레이어의 출력을 정규화.
- 드롭아웃 (Dropout): 학습 중 일부 뉴런을 랜덤으로 비활성화.

✓ 사용 모델 :

- 심층 신경망(Deep Neural Network, DNN) 기반의 이진 분류 모델 사용.
- 이진 분류 모델(Binary Classification)모델을 사용하여, 출력 레이어가 단일 뉴런을 가지고 있으며, 이 뉴런의 활성화 함수로 **sigmoid**를 사용함.

03 프로젝트 수행 결과

✓ 문제점 :

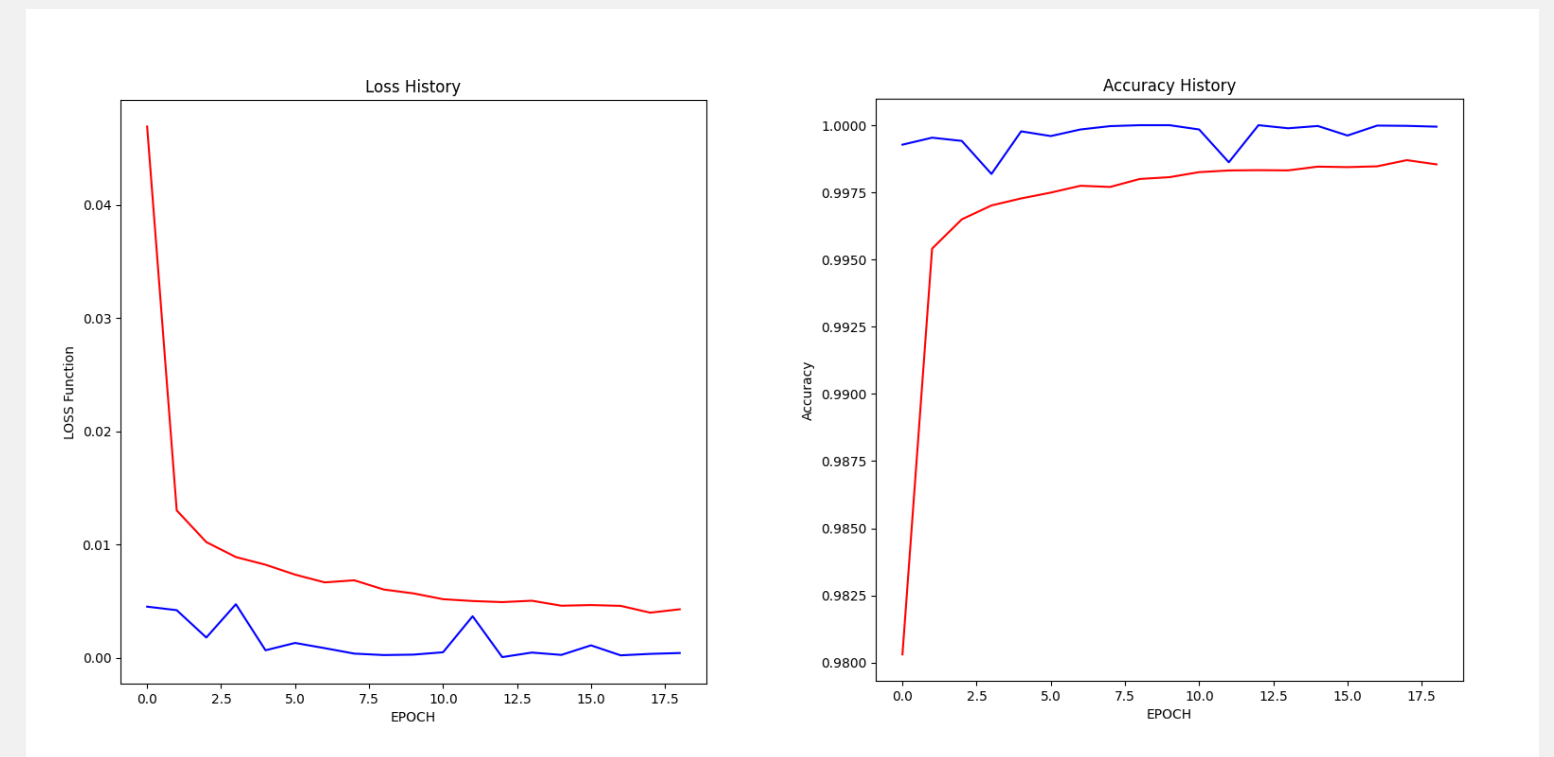
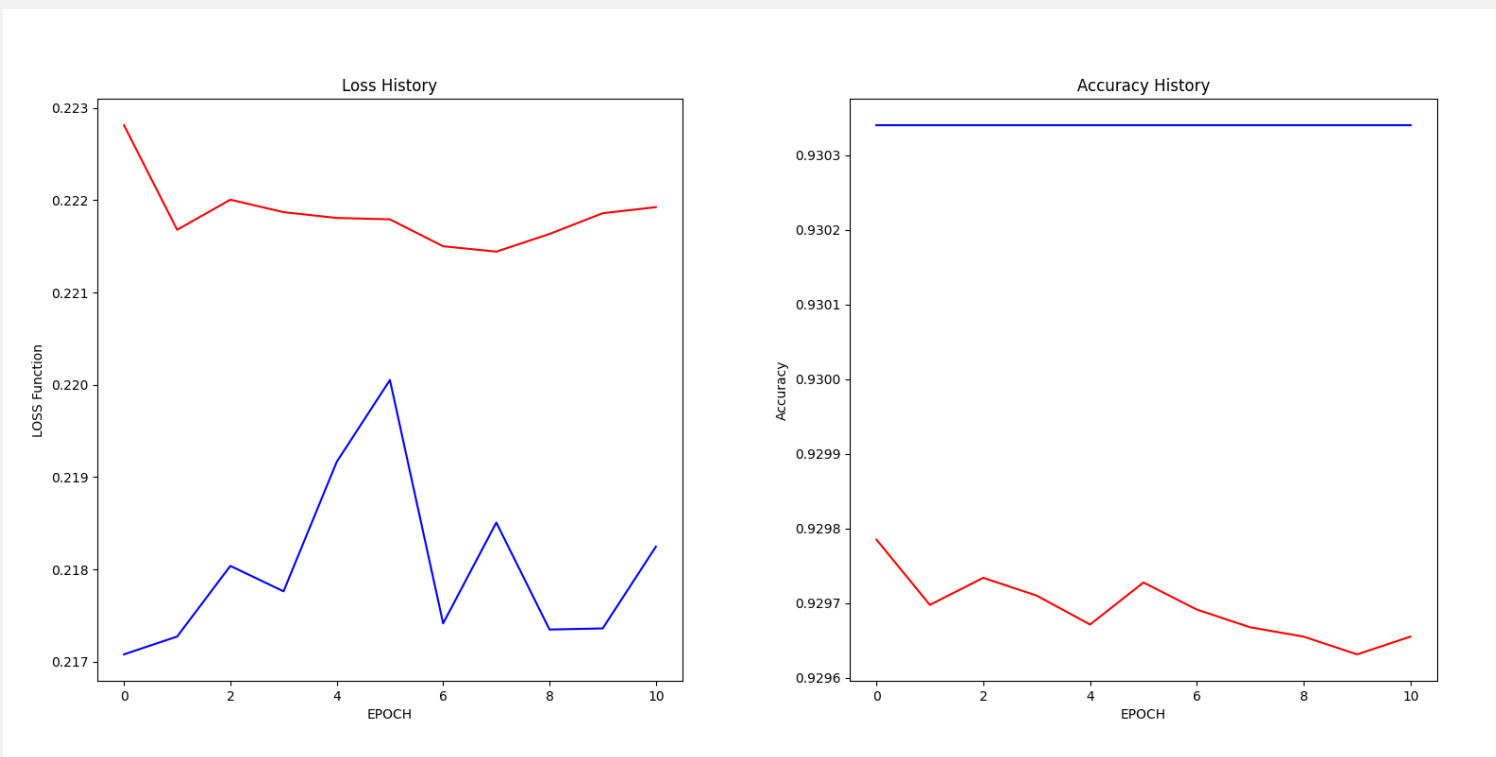
테스트 손실도, 테스트 정확도와 검증용 손실도, 검증용 정확도의 차이가 큼

✓ 원인 파악 :

데이터의 양이 백만개로 충분하여 원인이 과적합이 문제인 것 같음.

✓ 해결 방법 :

각 층을 정규화하는 배치정규화(Batch Normalization)와 일부 뉴런을 랜덤으로 비활성화 하는 드롭 아웃(Dropout)을 사용하여 일반화 성능을 높여 과적합을 방지



03 프로젝트 수행 결과

✓테스트 : 예측값 출력하기

```
***** TEST START *****
성별 : 남자, 연령대 : 70대이상, 총 콜레스테롤 : 132, 중성지방 : 168, 고밀도 콜레스테롤 40
심혈관 질환 위험도 : 67% (주의)

종료 코드 0(으)로 완료된 프로세스
```

```
***** TEST START *****
성별 : 여자, 연령대 : 20대, 총 콜레스테롤 : 110, 중성지방 : 80, 고밀도 콜레스테롤 50
심혈관 질환 위험도 : 24% (안전)

종료 코드 0(으)로 완료된 프로세스
```

심혈관 질환에 대한 AI 모델을 생성한 후
빅 데이터를 활용하여 AI모델을 학습 시켰다.

성별, 연령대, 총 콜레스테롤 수치, 중성지방 수치,
고밀도 콜레스테롤 수치를 X값으로 입력 하여
해당 AI모델 테스트 했다.

각 입력값에 따라 심혈관 질환 위험도는 다르게 나온 것을
확인 할 수 있었다.

03 프로젝트 수행 결과

✓ UI 페이지

나의 건강 데이터 입력

신장: 180

총 콜레스테롤: 320

몸무게: 80

중성지방: 280

공복혈당: 180

고밀도 콜레스테롤(HDL): 20

최고혈압: 200

☒ 비흡연

☐ 흡연

☐ 금연

☐ 비음주

☒ 음주

위험도 확인

Health Guard

미래 나의 건강은?

유시온 님 성별: 여성 연령: 27 세

심혈관 질환 위험도 54 %

뇌혈관 질환 위험도 87 %

고혈압 위험도 100 %

당뇨 위험도 99 %

메인으로

Health Guard

미래 나의 건강은?

유시온 님 성별: 여성 연령: 27 세

심혈관 질환 위험도 27 %

뇌혈관 질환 위험도 38 %

고혈압 위험도 10 %

당뇨 위험도 40 %

메인으로

Health Guard

위 사진은 데이터 입력 페이지와 각 질환 위험도를 표시해주는 페이지를 보여준다.

‘나의 건강 데이터 입력’ 페이지에서 각 데이터를 입력시 서버로 데이터가 전달이 된다. 서버에서는 AI파트로 데이터를 전송한 후 그 데이터에 대한 예측값을 받아 다시 ‘미래 나의 건강은?’ 페이지에 예측값을 그래프로 표시해준다.

04 자체 평가 의견

이번 프로젝트에서 텐서플로우(TensorFlow) 라이브러리를 사용하여 건강 검진 데이터 결과를 학습시키고, 당뇨 및 뇌·심혈관 질환에 대한 위험도를 예측하는 모델을 생성했습니다. 딥러닝을 학습하면서 주제에 따라 사용하는 모델, 활성화 함수 등이 각각 다르다는 것을 알게 되었습니다.

특히, 과적합 문제에 대해 신경을 썼습니다.과적합을 방지하기 위해 여러 방법이 있었는데 이 방법 또한 너무 많이 사용할 시 오히려 과적합이 발생할 수 있어 지속적으로 변경하면서 확인이 필요했습니다. 또한, 적절한 데이터를 찾는 것도 매우 중요하다는 것을 느꼈습니다. 주제를 정하고 데이터를 사용하여 모델을 학습시킬 때 예측값이 잘 나오지 않아서 데이터를 다시 확인해 보니 사용하기에 적합하지 않은 데이터여서 중간에 데이터를 변경하기도 했습니다.

비록 이번 처음 제가 계획한 대로 프로젝트를 완료했다고 할 수 없지만, 딥러닝에서 주의해야 할 부분들을 배우고 기본적인 개념, 코드적인 이해를 높이는 시간이 되었습니다. 딥러닝 AI 부분은 처음에는 어렵게 느껴졌지만, 이번 프로젝트를 통해 조금씩 다뤄볼 수 있었습니다.

감사합니다

Thank you