



한국정보과학회

KOREAN INSTITUTE OF INFORMATION SCIENTISTS AND ENGINEERS

http://www.kiise.or.kr

# Korea Software Congress 2020

“뉴노멀의 중심, 인간과 함께가는 소프트웨어!”

2020. 12. 21.(월) ~ 23.(수) / Online

12.21

12.22

12.23



주요행사

[Top Conference 세션]

- 차세대 시스템 세션
- 차세대 소프트웨어 세션
- 차세대 인공지능 세션
- 차세대 빅데이터 세션

[워크샵]

- 인공지능 클라우드 연구센터 자체워크숍
- 바이오 데이터 분석 분석 최신 기술 동향
- Korea Post-NeurIPS-2020 Workshop

[기초강연]

- 김창용 원장(NIPA)
- 최양희 교수(서울대, 과기부 전 장관)

[초청강연]

- 지성태 대표(Arte TV)
- 이주석 전무(인텔코리아)
- 김명철 교수(KAIST)
- 이광근 교수(서울대)

[박사포럼]

- 김박사, 어서와!

[Top 연구실 세션]

- KETI 세션
- ETRI 세션

[워크샵]

- 우수학술대회목록 개편위원회 보고회
- 컴퓨터 역사워크샵 - 인공지능
- UHD 방송환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 개발
- 뉴로모픽 컴퓨팅 SW 플랫폼 기술 발표회
- 빅데이터 엣지클라우드 서비스 워크샵
- 뉴노멀 시대의 인공지능 교육

[워크샵]

- 연구윤리워크샵
- 학회발전토론회

[경진대회]

- Intel Korea와 함께하는 사물인식+드론 경진대회

후원

GOLD

LINE

intel

kakao

kt

BRONZE

dcvb  
대구권발전위원회

cen  
아이티센  
ITcen

SK  
브로드밴드

SK  
hynix

130

We Believe  
INFRANICS

HYUNDAI

# 2020 한국소프트웨어종합학술대회(KSC2020)

2020년 12월 21일(월) ~ 23일(수), Online



|                                                             |                                         |      |
|-------------------------------------------------------------|-----------------------------------------|------|
| 501. 그래프 컨볼루션 네트워크와 유전자 발현 데이터를 활용한 약물 반응 예측 모델             | 김성훈 · 배석훈 · 박은화 · 조거리                   | 1466 |
| 502. 딥러닝 기반 Java 프로그램 자동 코딩 가이드 모델                          | 홍태하 · 우홍욱                               | 1469 |
| 503. 수정된 AE 모델을 이용한 X-ray 영상 개선 기법                          | 김유영 · 이 찬 · 정지훈 · 서효경 · 김호준             | 1472 |
| 504. COVID-19 전파 위험도 추정 : 물체 및 장면 감지를 사용한 실시간 화면 분석         | 공재호 · 이민호 · 이승민 · 지호진 · 최성원 · 신은섭 · 배성호 | 1475 |
| 505. 협력 로봇의 다중 위치 볼트 풀림 진단 SVM 축-동작 혼합 모델                   | 백지훈 · 유동연 · 박예슬 · 이정원                   | 1478 |
| 506. AITSC-NET: COVID19 진단을 위한 밝기보정, 객체 분할 및 분류 신경망         | 남주현 · 김수정 · 이상철                         | 1481 |
| 507. 온라인 서점 도서 데이터를 활용한 도서 주제 자동분류모델 구현                     | 김무성                                     | 1484 |
| 508. 딥러닝 기반 의료 설문 결측 데이터 보간                                 | 이선민 · 김동욱 · 이태현 · 김민태 · 김학재 · 이상주       | 1487 |
| 509. 말기신질환자의 마취 중 연속 혈압 예측 모델                               | 권도윤 · 박명석 · 박성아 · 우지영 · 강아름 · 정영섭 · 박선영 | 1489 |
| 510. OCR 기술을 이용한 반려동물 건강검진표 문자 판독                           | 정희영 · 홍헬렌                               | 1492 |
| 511. Ensemble 기법을 적용한 돌발성 난청 예측 예측 네트워크                     | 최홍규 · 김태현                               | 1495 |
| 512. 가우시안 커널 함수를 활용한 건강 보험료 예측 모델                           | 윤도원                                     | 1498 |
| 513. 운동 횟수와 속도 측정을 위한 딥러닝 기반의 수축 및 이완 분류                    | 김민영 · 김정화 · 정진우                         | 1501 |
| 514. 하이퍼레저 패브릭을 이용한 전자투표 모델의 설계 및 구현                        | 김성환 · 마지혜 · 성혜림 · 신유진 · 정경연 · 조강우 · 신상욱 | 1504 |
| 515. 블록체인 이더리움 기반 저작권 거래 서비스 DApp                           | 이현경 · 이지현 · 박재연 · 조은선                   | 1507 |
| 516. 디지털 이미지 저작권 보호를 위한 스크린 캡처 시스템                          | 박예린 · 이대호                               | 1510 |
| 517. WebAssembly 기능 도입으로 인해 크롬에서 발생한 오류들의 분석                | 이가현                                     | 1513 |
| 518. 개방형 OS에서 USB 개인정보 유출 방지 시스템                            | 현상엽 · 유준모 · 김규민 · 안종석                   | 1516 |
| 519. PX4 Autopilot을 활용한 드론 악성 소프트웨어 개발                      | 김성수 · 조진성                               | 1519 |
| 520. 개인건강기록의 프라이버시를 위한 블록체인 기반 영지식증명 보안 기술 설계               | 김유정 · 박서윤 · 김철수 · 이승현                   | 1522 |
| 521. Whitelist Filter를 통한 XSSDS의 개선방안                       | 강민우 · 정태명                               | 1524 |
| 522. Baremetal 펌웨어 보안 강화 기법                                 | 박우진 · 조진성                               | 1527 |
| 523. 드론의 RF 신호를 이용한 재사용 공격 방법                               | 이준오 · 조진성                               | 1530 |
| 524. TLS-암호화된 트래픽의 불변 특성 추출을 통한 멀웨어 패밀리 분류 방법의 가능성 분석       | 하준서 · 노희준                               | 1533 |
| 525. 수혈자 보호를 위한 블록체인 기반 전자 혈액 관리 체계 개발                      | 유안지 · 조상영                               | 1536 |
| 526. 로그 출력에 기반한 단일 연결된 비트코인 노드의 압축블록 전달 시간 측정 및 분석          | 김애리 · 맹수훈 · 에싸이드 매리암 · 주홍택              | 1539 |
| 527. 군집화 알고리즘을 이용한 무선 공격패킷 검출                               | 이경하 · 진영훈 · 서예진 · 정주용 · 정종민             | 1542 |
| 528. 뇌파 기반 드론 컨트롤을 위한 애플리케이션                                | 이승윤 · 강현석 · 김소람 · 김사온 · 안민규             | 1544 |
| 529. 계층 구조를 제공하는 P2P 기반 하이퍼레저 패브릭을 이용한 분산 파일 저장소            | 홍승환 · 이도영 · 박상원                         | 1547 |
| 530. 증강현실에서 적용되는 강화학습 상호작용 연구                               | 이승준 · 조상영                               | 1550 |
| 531. Haar Cascade Classifier와 LBPH 알고리즘 기반으로 객체 식별 모델 연구 사례 | 정경희 · 조성혁 · 김장환 · 김영철                   | 1553 |

# 2020 한국소프트웨어종합학술대회(KSC2020)

2020년 12월 21일(월) ~ 23일(수), Online



|                                                                                         |                                         |      |
|-----------------------------------------------------------------------------------------|-----------------------------------------|------|
| 440. 임베디드 장치에서의 고성능 추론을 위한 recursive한 레고 필터를 활용하여 CNN 모델을 학습하는 기법 연구 .....              | 정희수 · 신동균                               | 1283 |
| 441. Random CNN과 유전 알고리즘을 이용한 신경망 구조 .....                                              | 조한희 · 손은영 · 박희민                         | 1286 |
| 442. 스마트팩토리 로봇간 협력 프로토콜 구축을 위한 CNN 기반 작업 프로그램 인식 모델 .....                               | 김양곤 · 유동연 · 박예슬 · 이정원                   | 1289 |
| 443. 심층신경망의 불확실성 기반 열린 집합 도메인 적응 .....                                                  | 박경연 · 허재필                               | 1292 |
| 444. COVID19와 유행성 인플루엔자의 비교 예측 프로토타입 모델 .....                                           | 유승빈 · 김영희 · 김장환 · 김영철                   | 1295 |
| 445. 물체 지향적 메이즈: 일반화를 위한 강화 학습의 새로운 벤치 마크 환경 .....                                      | 이민후 · 김기범 · 이민수 · 장병탁                   | 1298 |
| 446. Auto Encoder를 이용한 위치데이터의 차원 확장 및 GAN학습 .....                                       | 송하운 · 전희준                               | 1301 |
| 447. AI분석을 통한 자동 밸런스 패치 제안 기술 연구 .....                                                  | 임용식 · 김재우 · 고재호 · 김준희 · 나용수             | 1304 |
| 448. Convolutional AutoEncoder를 이용한 GPS Data Denoising .....                            | 장준영 · 안태진 · 박희민                         | 1307 |
| 449. 세이버매트릭스 통계를 활용한 머신러닝 알고리즘 간의 성능 비교 .....                                           | 양윤석 · 김영훈                               | 1310 |
| 450. 상관관계 정규화와 동적 필터 가지치기를 이용한 심층 신경망 압축 .....                                          | 전희선 · 조인천 · 배성호                         | 1313 |
| 451. 변형가능 컨볼루션을 다양한 환경에 적용하기 위한 정적 계산 기법 .....                                          | 김상범 · 장하영                               | 1316 |
| 452. 딥러닝 기반 강화학습 모델 성능 비교 - 국내 주식시장 사례연구 .....                                          | 우메모토 세이야 · 황석형 · 김응희 · 김수환 · 송경환 · 김민경  | 1319 |
| 453. 한국인 안면 이미지를 활용한 마스크 착용 데이터 증강 및 마스크 착용 얼굴 인식 .....                                 | 류재희 · 정은지 · 김영민                         | 1322 |
| 454. YOLOv4를 이용한 총기 자동 검출 모델 생성 방법론 .....                                               | 박주원 · 김평강 · 김재원 · 홍참길                   | 1325 |
| 455. 전체 페이지 단위 한글 손글씨 인식을 위한 태스크 특정 데이터 증강 기법 .....                                     | 하남규 · 황지영 · 김혜인 · 이세진 · 구명완             | 1328 |
| 456. 관광지 소개 유튜브 동영상에서 딥러닝을 이용한 샷 단위 대표 음식 검출 방법 .....                                   | 나형운 · 장성욱 · 정민수 · 남종호                   | 1331 |
| 457. 배경 패딩 데이터 증강을 통해 축척에 강인한 딥러닝 기반 수화 인식기 .....                                       | 정성우 · 김민석 · 유준혁                         | 1334 |
| 458. 영상 콘텐츠 트렌드 분석/조화 및 추천 AI 시스템에 관한 연구 .....                                          | 권 준 · 박지홍 · 이경수 · 김 현                   | 1337 |
| 459. 이미지 내 시그니처 객체 식별을 통한 실내 위치 국지화 기법 .....                                            | 윤동희 · 박예슬 · 장현웅 · 이정원                   | 1340 |
| 460. FPN 과 Anchor 스케일 보정을 통한 드론 촬영 영상의 객체 탐지 성능 향상 .....                                | 김병철 · 김선후 · 조윤섭 · 이민지 · 박성식 · 송윤석       | 1343 |
| 461. AI 기반 얼굴 위치 가이드 및 Style Transfer 모바일 어플리케이션 .....                                  | 김민경 · 홍헬렌                               | 1346 |
| 462. 지능형 웹 서비스를 위한 물체 검출, 색상 추정과 위치를 이용한 영상 묘사 .....                                    | 옥수빈 · 이대호                               | 1349 |
| 463. 초광대역 무선 신호 기반 사람의 자세 추정 기법 .....                                                   | 임수빈 · 김승현 · 채근홍 · 강영훈 · 손하영 · 최계원 · 김유성 | 1352 |
| 464. 저해상도 이미지 기반 손실 함수를 활용한 해상도 개선 .....                                                | 류재현 · 허재필                               | 1355 |
| 465. Data Augmentation을 사용한 Image Classifier Training의 Input Pipeline Overhead 분석 ..... | 이하연 · 이계원 · 전병곤                         | 1358 |
| 466. 스마트폰으로 촬영된 알약 이미지의 식별 마크 인식 모델을 이용한 알약 인식 및 정보 제공 시스템 .....                        | 배소현 · 손수현 · 전예진 · 황선정                   | 1361 |
| 467. Faster R-CNN을 이용한 축구 경기 하이라이트 영상 추출 시스템 .....                                      | 최현진 · 정현구 · 최재영 · 정윤 · 김병만              | 1364 |



# COVID19와 유행성 인플루엔자의 비교 예측 프로토타입 모델

유승빈<sup>01</sup> 김영희<sup>2</sup> 김장환<sup>3</sup> 김영철<sup>4</sup>

<sup>1,2</sup>홍익대학교 컴퓨터정보통신공학과

<sup>3,4</sup>홍익대학교 소프트웨어 공학 연구실

usb7121@gmail.com, cin05189@naver.com, janghwan@selab.hongik.ac.kr, bob@hongik.ac.kr

## A Prototype Model for Comparison and Prediction Between COVID19 and Influenza in Korea

Seung bin Yoo<sup>01</sup> Young hee Kim<sup>2</sup> Jang hwan Kim<sup>3</sup> R. Young Chul Kim<sup>4</sup>

<sup>1,2</sup>Dept. of Computer Information and Communication Engineering, Hongik Univ.

<sup>3,4</sup>Software Engineering Lab, Hongik Univ.

### 요 약

전 세계적인 대유행과 사회적인 격리로 막대한 경제적, 정신적, 육체적인 문제들이 발생하고 있다. 현재 대안인 의료계의 백신, 의공학계의 다중검사도구 연구가 진행 중이다. 또한 앞으로 트윈데믹 상황에 데이터 기반 소프트웨어 예측 모델 및 도구를 통해 비교 및 예측 방법도 필요하다. 이를 위해, 코로나 바이러스와 한국형 독감바이러스의 대한 식별 예측 프로토타입 모델을 제안한다. 즉, 확률형 머신러닝 알고리즘을 적용하여 데이터수집 및 분석을 통해 의료 관련 검사 및 진단도구의 사용 없이도, 독감과 코로나 바이러스를 비교 예측이 가능하기를 기대한다. 실제 익명의 코로나 확진자 데이터와 질병관리청 독감 증상정보 기반 가상 독감환자 데이터를 예측모델에 적용하여 실험하였다. 각각 두 가지 경우 모두 올바르게 식별하였다. 원하는 기댓값을 얻었으며 이후 보다 향상된 예측 모델을 구성하고자 한다.

### 1. 서 론

현재의 지속적인 문제인 코로나 바이러스(COVID-19)가 전 세계로 확산되면서 막대한 경제적, 정신적, 육체적 문제들이 발생하고 있다. 이에 세계보건기구(WHO)는 2020년 3월 팬데믹(Pandemic)을 선언하면서 전염병의 위험을 경고하였다. 이런 상황인 가운데 일교차가 커짐에 따라 면역력이 급격히 떨어지면서 독감과 코로나 바이러스가 동시에 다가오는 '트윈데믹(Twindemic)'에 대한 우려가 커지고 있다.

세계보건기구에 따르면, 이들은 접촉과 비말(침방울)에 의한 감염 경로, 고열, 피로, 기침을 유발하는 호흡기 질환 등에 공통점이 있다. 차이점으로는 전파속도, 전염력, 2차 감염, 치료법, 치명성 등이 있다[1,2]. 정부에서는 이런 문제에 대비하고, 국내외 연구진들이 한 번에 진단 가능한 키트를 개발하는 중이다. 현재 일반인이 위에서 언급한 두 질병을 구분하는 것은 매우 어렵다. 본 논문에서는 데이터 기반의 비교 예측 프로토타입 모델을 제안한다. 이는 확률형 머신러닝 알고리즘 기반 데이터를 수집하고 분석함으로써 코로나 바이러스와 한국형 독감 바이러스를 예측한다. 이 방법을 통해 일반 사람들이 검사도구의 사용 없이도 두 질병을 보다 쉽게 식별하기를 기대할 수 있다. 본 논문의 구성은 다음과 같다. 2장에서는 확률형 머신러닝 알고리즘을 소개한다. 3장에서는 비교 예측 프로토타입 모델에 대해 설명한다. 4장에서는 기존 모델과 다른 알고리즘을 적용하여 비교하는 실험이다. 본 연구에서 모델에 환자 데이터를 적용해 실험한 결과 코로나 확진자 및 독감 확진자 식별에 성공하였다. 5장은 결론 및 향후 연구를 언급한다.

### 2. 관련 연구

#### 2.1 Naive Bayes

Naive Bayes는 조건부 확률 모델로 각 사건들을 독립으로 가정한다. 그를 통해 베이즈 정리에 근거한 발생확률을 기반으로

로 분류한다[3]. 이 모델의 장점은 일부 확률 모델에서 지도학습 환경에 매우 효율적으로 훈련될 수 있으며, 필요한 파라미터를 추정하기 위한 트레이닝 데이터양이 매우 적다. 우리는 수집한 데이터를 비교 분석하여 예측된 결과를 확률로 나타내기 위해 제안한 예측프로토타입 모델에 적용한다.

#### 2.2 최근접 이웃 분류(K-Nearest Neighbors Classifier)

최근접 이웃 분류 모델은(KNN) 입력값에 대한 결과값을 예측할 때 훈련 데이터셋에서 가장 가까운 '최근접 이웃'을 찾아 예측에 사용한다[4]. 입력한 속성값과 가장 거리가 가까운 k개의 관측치를 통해 결과를 예측할 수 있고 많이 조정할 필요 없이 좋은 성능을 발휘한다.

### 3. 비교 예측 프로토타입 모델

#### 3.1 개발개요 및 데이터 전처리

코로나와 독감 바이러스를 비교하여 바이러스별 증상의 차이를 발견하고 사용자의 증상이 어떤 바이러스에 확진될 확률이 높은지 예측하는 두 가지 모델을 만든다. 그리고 각 모델의 차이 및 정확도를 비교한다. 코로나 바이러스와 비교할 독감 바이러스는 올해 대중적으로 유행하고 있으며 가장 유행의 빈도가 높은 'A형 독감'이다. 코로나 바이러스 환자와 독감환자들의 증상데이터를 머신러닝을 통해 학습시키고 2가지 다른 모델을 통해 타겟 데이터를 어떻게 분류하는지 비교한다.

먼저 코로나 바이러스와 독감바이러스에 확진된 환자의 증상이 담긴 데이터를 수집하고[5] 전처리 작업을 통해 정제한다. 여러 데이터 속성 값에는 발열을 비롯한 7개의 다른 종류의 증상들과 Age, Gender, Severity 등의 식별 데이터로 구성된다. 머신러닝이나 딥러닝 알고리즘은 수치로 된 데이터만 이해할 수 있기 때문에 Age, Gender, Severity를 범주형 데이터에서 boolean 데이터타입으로 변환하였다. Gender의 데이터 값은 Female, Male, Transgender를 사용한다. 이 3가지 데이터값을

컬럼으로 추가하여 G\_Female, G\_Male, G\_Transgender로 변환한다. 변환은 해당 컬럼에만 1을 표시하고, 1을 제외한 컬럼은 0으로 표기하는 원-핫 인코딩 방법을 이용하였다. 그림 1은 원-핫 인코딩 변환 방법이다.

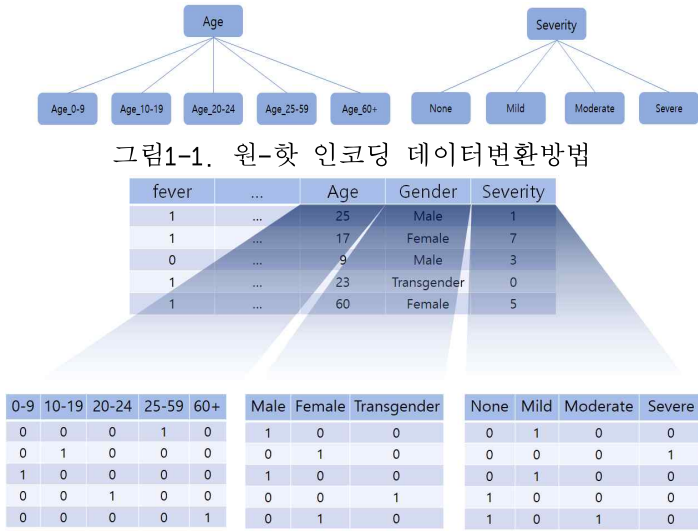


그림1-2. 원-핫 인코딩 데이터변환방법 테이블

A형 독감의 각 증상이 담긴 데이터를 수집하고[6], 유병률을 기준으로 독감 증상 데이터를 증상의 단위와 정도에 따라 테스트 데이터로 생성했다. 코로나 바이러스 데이터와 마찬가지로, A형 독감 바이러스 증상 속성 값을 Fever, Cough, Difficulty-in-Breathing, Sore-Throat, Nasal-Congestion, Diarrhea, Headache으로 통일했다. 독감 증상 데이터는 코로나 바이러스 데이터와 동일한 크기로 생성하고, 데이터의 학습을 위해 두 데이터를 결합했다.

```
x_data = data.values[:, 0:19] # symptoms & info
y_data = data.values[:, 19] # Result COVID-19, Influenza
```

그림 2. 데이터 테이블 전처리 코드

그림 2는 데이터 테이블의 전처리 코드이다. 그림 2의 코드에서 x\_data에 코로나 바이러스와 독감증상 속성들을 저장하고 결과 데이터를 y\_data에 담았다.

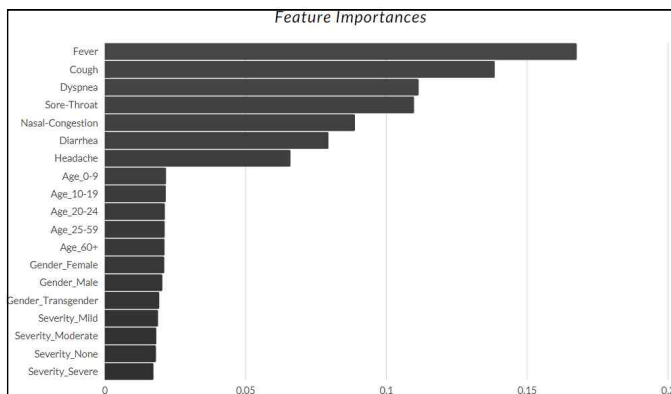


그림 3. 데이터 속성값의 중요도

그림 3은 데이터 전처리 과정을 적용한 이후 구성한 모델이 예측함에 있어 각각의 속성값에 대해 영향 받는 정도를 수치화하여 그래프로 나타낸 것이다.

### 3.3 기존 Naive Bayes를 적용한 비교 예측 프로토타입 모델

데이터 분류 알고리즘으로 Naive Bayes 알고리즘을 이용하였다. 이 알고리즘의 이벤트 모델은 Gaussian, Multinomial, Bernoulli등으로 데이터의 종류와 특성에 따라 다양한 이벤트 모델들이 존재한다[7]. 수집한 데이터는 원-핫인코딩을 통해 이진속성으로 데이터가 정제 되어있으므로 여러 이벤트 모델 중 BernoulliNB 모델을 사용하였다. 또한 각 모델의 성능 비교를 위해 학습 데이터를 훈련시키고 학습 데이터가 훈련이 잘 되었는지 검증 데이터를 통해서 모델 성능을 평가한다. 구체적인 평가 방법은 검증 데이터셋 y(결과값)를 활용하여 검증 데이터셋 x(속성값)를 predict함수에 파라미터로 넣어준다. 그러면 기존의 학습된 데이터 y값과 대비 정확도를 확률로 나타낸다.

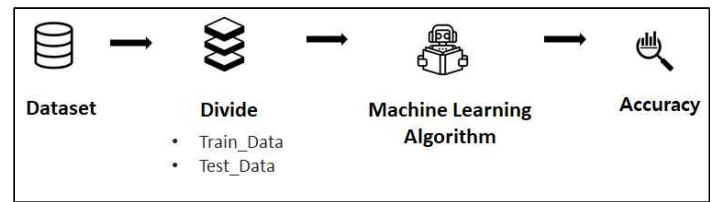


그림4. 기존 모델 구성 방식

그림 4는 기본적인 모델 구성 방식이다. 데이터 세트를 학습 데이터, 테스트 데이터로 구분하고 지도학습 알고리즘에 적용하여 모델을 구성한다. 학습 데이터와 테스트 데이터의 분리 및 알고리즘 적용 과정에서 argument 값을 default(None)로 할당하였다. 이후 모델 성능을 측정한 결과 약 70%에서 80%의 정확도를 나타냈다. 이후 정확도 향상을 위한 최적의 argument값을 할당하기 위해 테스트과정을 거쳤다.

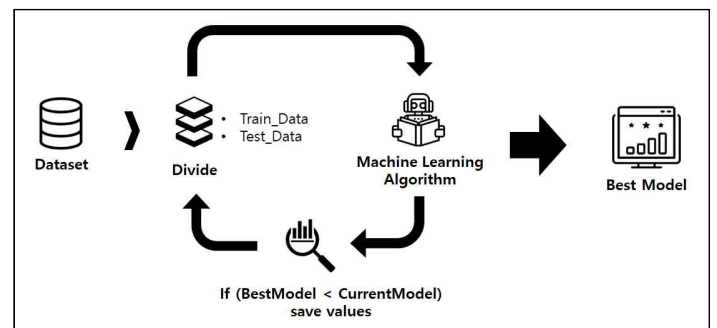


그림5. 개선한 모델 구성 방식

그림 5는 최적 학습 및 개선한 모델 구성 방식이다. 우리는 모델의 성능향상을 위해 학습시킬 데이터와 테스트용 데이터를 구분하는 것과 지도학습 알고리즘을 적용하여 학습하는 과정을 반복했다. 테스트 데이터 세트의 비율인 test\_size 값을 0.001에서 0.499까지 할당했고, 이를 위한 시드값 random\_state를 1에서 100까지 할당했다. 알고리즘 내 파라미터 값에 따른 성능차이를 분석하고 개선되는 방식으로 그림과 같이 반복 학습하여 최적의 모델을 구성하였다.

```
print(" score : %f" % model.score(x_test, y_test))
score : 0.937500
```

그림6. 최적화 모델 성능테스트 결과값

그림 6은 최적화 모델 성능테스트의 결과값이다. 성능 측정결과 약 88%에서 약 94%로 상승하였다.

우리는 최적화된 데이터셋을 모델에 학습시켰고, 두 개의 환자 데이터를 모델에 적용하였다. 첫 번째 데이터는 전라남도 광주 지역의 익명 확진자에게 동의를 얻어 데이터를 수집하였다. 수집한 데이터는 발열, 기침, 호흡곤란, 코막힘 증상 및 증상의 심각도는 보통인 30대 남성이다. 두 번째 데이터는 질병관리청 독감 증상정보를 기반으로 발열, 기침, 인후통, 두통 증상 및 증상의 심각도는 심각한 20대 여성 환자 데이터를 사용했다[8].

```
COVID19_Confirmed_case = np.array([
    1, # Fever
    1, # Cough
    1, # Difficulty-in-Breathing
    0, 1, # Nasal-Congestion
    0, 0, 0, 0, 1, # Age_25-59
    0, 0, 1, # Gender_Male
    0, 0, 1, # Severity_Moderate
    0, 0])
Influenza_Confirmed_case = np.array([
    1, # Fever
    1, # Cough
    0, 1, # Sore-Throat
    0, 0, 1, # Headache
    0, 0, 0, # Age_20-24
    0, 0, 1, # Gender_Female
    0, 0, 0, 0, 0, 1 # Severity_Severe
])
```

그림 7. 환자 데이터

그림 7은 환자 데이터이다. 두 타겟 데이터를 predict 함수를 이용하여 분류된 데이터와 나이브 베이즈 알고리즘으로 학습된 데이터 사이의 유사도를 판단하여 각 환자 데이터가 코로나 바이러스에 가까운지, A형 독감에 가까운지 확률로 예측했다. 0 일 확률은 독감을 나타내고, 1일 확률은 코로나를 의미한다. 그림 8은 타겟 데이터 결과이다.

```
model.predict_proba(COVID19_Confirmed_case)
[[0.86730965 0.13269035]]
model.predict_proba(Influenza_Confirmed_case)
[[0.0792524 0.9207476]]
```

그림 8. 타겟 데이터 결과 (나이브 베이즈 모델)

그림 8에서 1열은 데이터 값이 코로나 바이러스일 확률이며, 2열은 독감일 확률이다. 첫 번째 환자 데이터는 환자의 문진에 비추어볼 때, 86.7% 확률로 코로나 바이러스, 13.3%의 확률로 독감 바이러스 진단에 양성 판정을 받을 수 있다. 두 번째 환자 데이터는 환자의 문진으로 비추어 볼 때, 7.9% 확률로 코로나 바이러스, 92.1% 확률로 독감 진단에 양성 판정을 받을 수 있다.

#### 4 Other Classifier 이용한 예측 모델

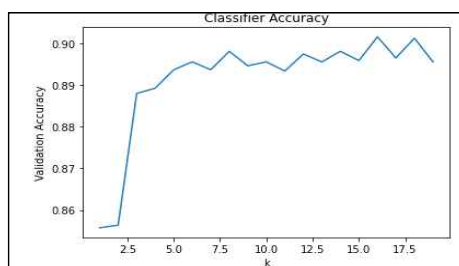


그림 9. KNN모델 성능측정 결과 그래프

동일한 실험을 다른 알고리즘에 적용해서 성능 변화와 정확도를 측정한다. 지도학습 알고리즘인 KNN 모델의 분류기법을 통해 동일한 실험을 수행했다. 그림 9과 같이 성능측정결과 k값이 16일 때 가장 높은 성능을 보인다. 즉, '최근접 이웃'을 16개를 찾아 학습하여 분류된 데이터들과 환자 데이터 사이의 유사한 정도를 판단하여 각 환자가 코로나 바이러스와 A형 독감 바이러스의 확진 확률을 예측하였다.

```
model.predict_proba(COVID19_Confirmed_case)
[[1. 0.]]
model.predict_proba(Influenza_Confirmed_case)
[[0. 1.]]
```

그림 10. 최근접 이웃 타겟 데이터 결과

그림 10은 최근접 이웃 학습 데이터 결과이다. 1열은 데이터 값이 코로나 바이러스일 확률이며, 2열은 독감일 확률이다. 실험한 두 가지 알고리즘 기반 예측모델 모두 식별에 성공하였다.

#### 5. 결론 및 향후 연구

본 논문은 코로나 및 독감 바이러스의 비교 식별 예측 프로토타입 모델을 제안한다. 이미 코로나 바이러스로 인해 팬데믹 한 상황에 있고 트윈데믹에 대한 우려가 커지는 가운데, 제안한 모델은 확률형 머신러닝 알고리즘을 적용하여 독감과 코로나 바이러스를 비교 예측한다. 본 연구를 통해, 일반 사람들도 보다 쉽게 검사도구 없이도 코로나 및 독감 바이러스를 식별할 것으로 기대한다. 이에 일정 수치 이상의 결과를 받은 사용자에게 코로나 진단 검사인 RT-PCR 등 검사를 권고하는 등 선제 조치가 가능할 것이다. 향후, 데이터 수집 및 관련 정보를 지속적으로 수집하여 좀 더 다양하고 실제 환자 데이터에 정확히 일치하는 속성을 추가할 것이다. 코로나 바이러스 및 독감의 다양한 증상, 발현 순서 등 다양한 종류의 데이터를 수집하고 지속적으로 모델을 학습시켜 성능이 개선되기를 기대한다. 코로나 바이러스, 독감 그 외 다른 질병에 대한 신뢰성 있는 데이터가 제공된다면 이 또한 적용 가능할 것이다. 이후 제안한 예측 모델의 서비스 제공을 위한 어플리케이션을 개발할 것이다. 이에 따라 코로나 바이러스 감염 진단 이전에 환자를 식별하여 전염성 질병에 대한 예방적 개입이 가능할 것으로 기대한다.

#### 참 고 문 헌

- [1] 코로나바이러스감염증-19, <http://ncov.mohw.go.kr>
- [2] MIT Technology Review, "These are 6 of the main differences between flu and coronavirus", Boitechnology Mar. 10. <https://www.technologyreview.com/2020/03/10/905405/flu-vs-coronavirus-6-differences/>
- [3] 염하늘, 방재근, 황명권, 황미영, 정한민, "나이브 베이즈 모델에 기반한 웹 뉴스 기사의 제목 추출", 한국정보과학회 학술발표논문집, pp. 479-481, 2014.12.
- [4] Jin Hyub Lee, Je Seoung Hong, R. Young Chul Kim, Hyun Seung Son, "Analysis of Integrated Renewable Energy Monitoring System Data using KNN for Pre-Processing", Asia Pacific Society for Computing and Information Technology, pp. 48-50, 2017.
- [5] COVID-19 Symptoms Checker, <https://www.kaggle.com/iamhungundji/covid19-symptoms-checker>
- [6] Ali Mohamed Zaki et al., "Post-pandemic influenza A (H1N1) virus detection by real-time PCR and virus isolation", Korean Journal of Microbiology, Vol. 55, No. 1, pp. 25-32, 2015.
- [7] 김동권, "나이브 베이즈 분류", 대한설비공학회 설비저널, Vol. 48, No. 12, pp. 94-96, 2019.12.
- [8] 질병관리청 인플루엔자(Influenza) 질병 정보, <http://www.kdca.go.kr/contents.es?mid=a21106010000>