

# 포트폴리오

-논문 및 프로젝트 경험-

유성민

# 목차

## 1. 자기소개

## 2. 논문 및 프로젝트 경험

- 1) 정수형 이질적 p차 자기회귀 모형의 IPO 수 예측 분석 (2022.03-2023.03)
- 2) 빅데이터 분석을 활용한 비즈니스 문제 해결 (2020.09-2020.11)
- 3) 농아인들을 위한 양방향 수어 번역기 웹사이트 개발 (2020.08-2020.09)
- 4) 빅데이터 기반의 지능정보 시스템 개발 (2020.06-2020.09)
- 5) 뉴스 빅데이터를 활용한 시각화 및 자연어 처리 (2019.11-2019.12)

# 1. 자기소개



**성명: 유성민**

**생년월일: 1996.11 (만 26세)**

**학력: 가천대학교 일반대학원 (4.38/4.5)**

**가천대학교 응용통계학과 (4.05/4.5)**

**주소: 서울 동작구 장승배기로**

# 1. 자기소개

저는 \_\_\_\_\_ 하는 지원자 입니다.

1. 보기 좋은 코드를 고민하는
2. 문제 해결을 끝까지 완수하는
3. 다른 직무에도 관심을 갖는
4. 소통을 좋아하는
5. 하는 일이 조직에서 어떤 영향력을 갖는지 고민하는

# 1. 자기소개

## 보유 기술

### [통계분석 툴]

R - 군집분석, 범주형 자료분석

SAS - 회귀분석, 다변량 분석

SPSS - 분산분석

### [개발 툴]

Python - 데이터 마이닝, 이미지/자연어/음성 분석

[웹/서버/DB 툴] - HTML5, CSS, Docker, SQLite

[인포그래픽 툴] - Tableau, Qlik

[협업 툴] - SLACK, GitHub (<https://github.com/YUSEONGMIN>)

## 2. 논문 및 프로젝트 경험

### 1) 정수형 이질적 p차 자기회귀 모형의 IPO 수 예측 분석

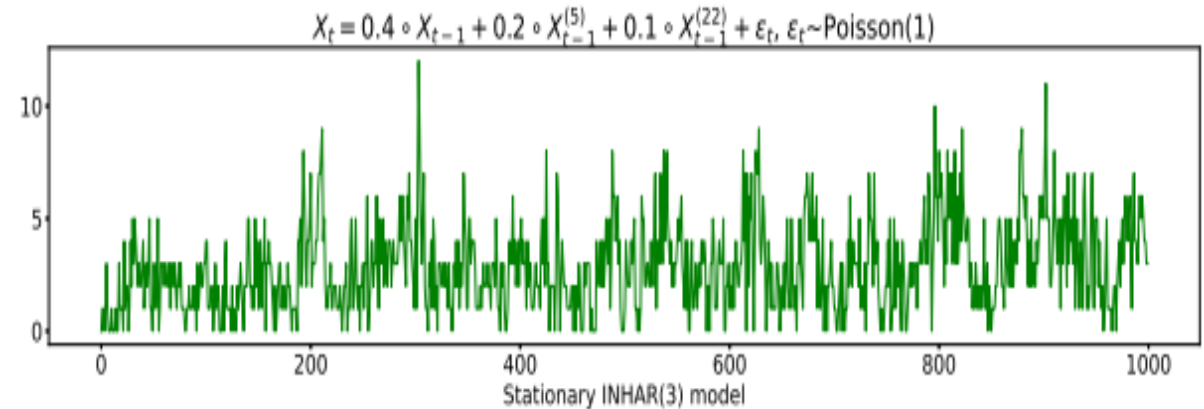
Communications for Statistical Applications and Methods  
2023, Vol. 30, No. 3, 273–289

<https://doi.org/10.29220/CSAM.2023.30.3.273>  
Print ISSN 2287-7843 / Online ISSN 2383-4757

#### Integer-Valued HAR( $p$ ) model with Poisson distribution for forecasting IPO volumes

SeongMin Yu<sup>a</sup>, Eunju Hwang<sup>1, a</sup>

<sup>a</sup>Department of Applied Statistics, Gachon University, Korea



목표: 정수형 이질적 데이터를 잘 반영하는 모형 구현

필요 역량: 수식에 대한 이해와 수학적 알고리즘 구현 능력

문제 발생: 실제 데이터가 모형에 맞지 않음

## 2. 논문 및 프로젝트 경험

### 1) 정수형 이질적 p차 자기회귀 모형의 IPO 수 예측 분석

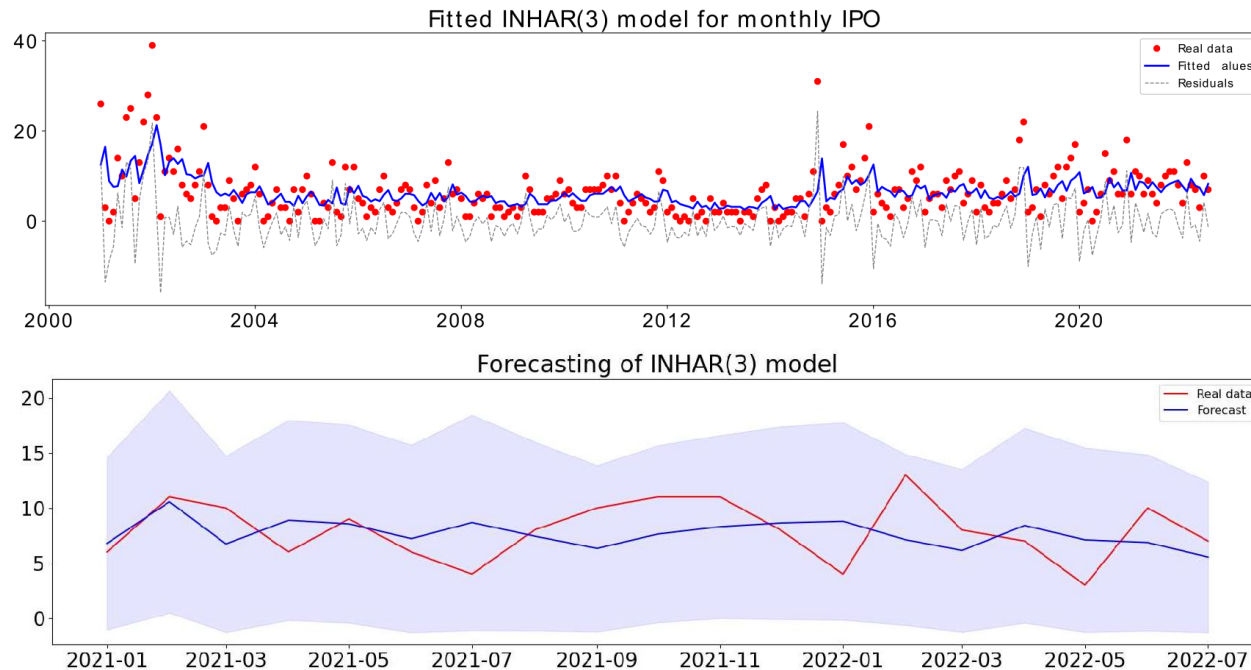
문제 해결 과정: 발생 가능한 문제 원인 파악 후 해결 방법 적용

- 1) 데이터가 이질적인 특성을 띄지 않는 경우
  - 이질적인 데이터를 다룬 논문을 참고하여 관련 데이터 수집
- 2) 모형 이론 조건을 만족하지 않는 경우
  - 정상성을 맞추기 위한 로그 변환 등 전처리 작업 수행

결과: 데이터가 모형에 잘 적합하고 비교 모형보다 성능이 좋게 나왔음

## 2. 논문 및 프로젝트 경험

### 1) 정수형 이질적 p차 자기회귀 모형의 IPO 수 예측 분석



배운 점:

- 1) 문제가 발생했을 때 포기하지 않고 방법을 찾고자 노력한다면 해결할 수 있다는 점과
- 2) 설계를 잘 하면 문제 발생을 줄일 수 있어 시간을 절약할 수 있다는 것을 배웠습니다.



## 2. 논문 및 프로젝트 경험

### 2) 빅데이터 분석을 활용한 비즈니스 문제 해결



0 5 10 15 20 25 30		0 5 10 15 20 25 30		0 5 10 15 20 25 30																
정말	너무	진짜	맛있어요	여기	다	맛도	완전	그냥	또	...	차라리	시켜먹었는데	왔습니 다	이건	맛있다고	반찬	싫네요	어떻게	Id	label
0	0	0	0	0	0	1	0	1	0	...	0	0	0	0	0	0	0	0	1	1
1	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	2	0
2	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	0

목표: 음식점의 고객 리뷰에서 음식 사진 분류와 부정적 댓글 필터링하기

필요 역량: 이미지 다중 분류와 감성 분석 능력과 다양한 모델에 대한 지식

문제 발생: 모델 성능이 좋지 않음

## 2. 논문 및 프로젝트 경험

### 2) 빅데이터 분석을 활용한 비즈니스 문제 해결

문제 해결 과정:

- 1) EDA를 통해 데이터 특징 파악 후, 적절한 전처리 작업
  - 긍정과 부정에서 공통으로 많이 등장한 단어들 제거
- 2) 다양한 모델 적용 및 하이퍼 파라미터 튜닝
  - 분류 분석에 사용되는 Random Forest나 이를 보완한 XGBoost 등 사용

결과: 이전보다 모델 성능을 올릴 수 있었고,  
감성 분석의 경우, 긍정에서는 “맛있어요”, 부정에서는 “다신” 중요도가 높게 나옴

## 2. 논문 및 프로젝트 경험

### 2) 빅데이터 분석을 활용한 비즈니스 문제 해결

```
1 print("Logistic의 가장 좋은 성능은 ",estimator_results[0].best_score_)
2 print("KNN의 가장 좋은 성능은 ",estimator_results[1].best_score_)
3 print("RandomForest의 가장 좋은 성능은 ",estimator_results[2].best_score_)
4 print("XGBoost의 가장 좋은 성능은 ",estimator_results[3].best_score_)
```

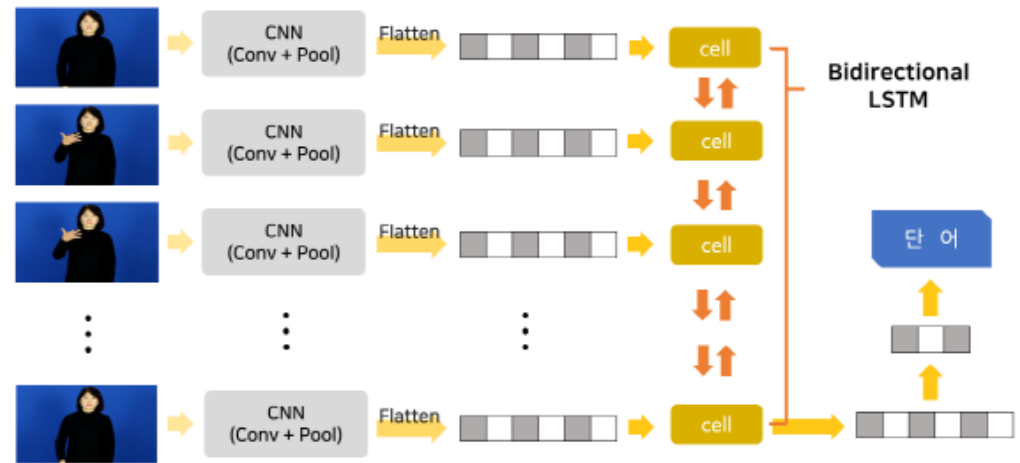
Logistic의 가장 좋은 성능은 0.7896731664050542  
KNN의 가장 좋은 성능은 0.6977328979655791  
RandomForest의 가장 좋은 성능은 0.7089497172627791  
XGBoost의 가장 좋은 성능은 0.7854408098511747

배운 점:

- 1) 왜 이런 분석을 했는지에 대해 아는 것이 중요하다는 것과
- 2) 리뷰 데이터 분석으로 마케팅에 어떻게 활용되면 좋을지 인사이트를 배웠습니다.

## 2. 논문 및 프로젝트 경험

### 3) 농아인들을 위한 양방향 수어 번역기 웹사이트 개발



목표: 수어 영상과 문장을 서로 번역하는 사이트 구현

필요 역량: 영상 및 자연어 처리, 웹 개발, DB 관리 능력

문제 발생: 동작에 맞는 단어가 나오지 않음

## 2. 논문 및 프로젝트 경험

### 3) 농아인들을 위한 양방향 수어 번역기 웹사이트 개발

문제 해결 과정:

#### 1) 데이터 추가

- 직접 영상을 찍어서 데이터 추가
- Augmentation을 통해 데이터 추가

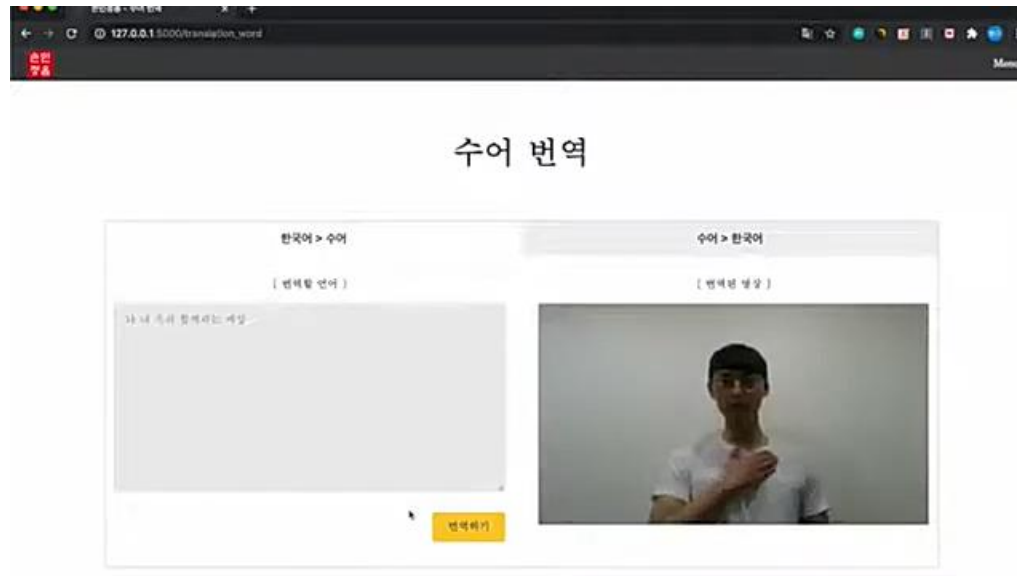
#### 2) 하이퍼 파라미터 조정

#### 3) 최후의 수단은 해당 단어 제외

결과: 전보다 더 많이 동작에 맞는 단어를 출력함

## 2. 논문 및 프로젝트 경험

### 3) 농아인들을 위한 양방향 수어 번역기 웹사이트 개발



배운 점:

- 1) 데이터 분석, 웹 개발, DB 구현 등 역할이 다양할 땐 더욱 더 소통이 중요하다는 것과
- 2) 조직 내에서 프로젝트를 할 때 작은 일은 없다는 것을 배웠습니다.

## 2. 논문 및 프로젝트 경험

### 4) 빅데이터 기반의 지능정보 시스템 개발



목표: 기계학습/딥러닝을 활용한 빅데이터 파이프라인 구축

필요 역량

- 파이썬 고급 문법: 자료구조, 디버그, 함수적 기법, 디자인 패턴
- 빅데이터 수집 및 전처리: SQL, 크롤링, NLP, Elastic Search 활용
- 기계학습: EDA, 앙상블 기법, 시계열/위치 기반 분석
- 딥러닝(강화학습): RNN, LSTM, GRU, DCGAN 알고리즘, Attention 기법
- 지식정보 시스템 개발: 파이프라인 구축, 웹 프로그래밍

## 2. 논문 및 프로젝트 경험

### 4) 빅데이터 기반의 지능정보 시스템 개발

배운 점:

- 1) Class, Attribute, Instance, Method의 개념
- 2) @property 데코레이터를 이용한 getter/setter 구현
- 3) 상속과 위임(컴포지션)의 차이
- 4) SQLAlchemy와 MongoDB를 이용한 DB 연동
- 5) KoNLPy와 NLTK를 이용한 Tokenizing 기법과 BPE(Byte Pair Encoding)
- 6) Zipf's Law와 Heap's Law 이론을 통한 TF-IDF 함수 생성
- 7) Keras와 Pytorch를 활용한 GAN 알고리즘 구현
- 8) 사이킷런의 GridSearchCV를 이용한 파이프라인 개발
- 9) VotingClassifier, Bagging, AdaBoostClassifier, Stacking의 차이



## 2. 논문 및 프로젝트 경험

### 5) 뉴스 빅데이터를 활용한 시각화 및 자연어 처리



**목표: R을 이용한 뉴스 빅데이터 분석 및 시각화**

**필요 역량**

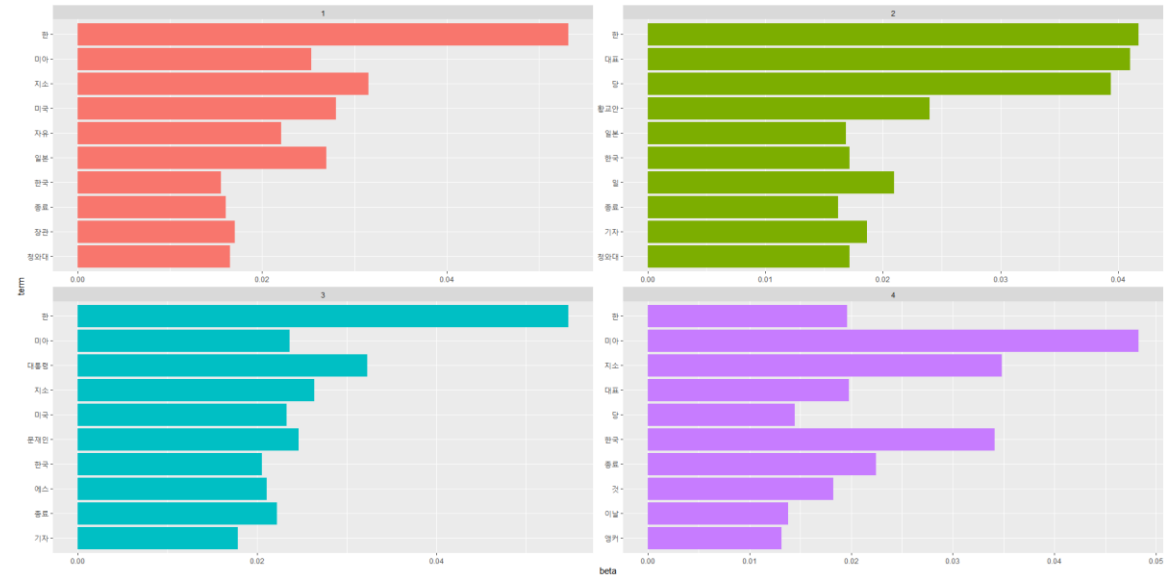
- 뉴스 기사라는 비정형 데이터의 이해
- R을 이용한 데이터 크롤링 / 텍스트 분석 / 시각화 기법

## 2. 논문 및 프로젝트 경험

### 5) 뉴스 빅데이터를 활용한 시각화 및 자연어 처리

배운 점:

- 1) KoNLP 패키지를 이용한 비정형 데이터 분석
- 2) tidyverse와 dplyr을 이용한 데이터 전처리
- 3) ggplot2를 활용한 다양한 시각화
- 4) TDM을 통해 계산된 TF-IDF 값으로 단어별 가중치 부여
- 5) Apriori 알고리즘을 이용한 연관규칙 분석
- 6) Latent Dirichlet Allocation(LDA) 기법을 적용한 토픽 모델링 생성



저의 강점을 통해 회사에 기여하고  
열심히 배우며 발전하겠습니다.  
감사합니다.