

수도권 도시 환경 요인에 따른 범죄 발생 건수 예측

장예원*, 김예림**, 박시현^o, 이재영**, 문유진**

*한국외국어대학교 이탈리아어통번역학과,

**한국외국어대학교 Global Business & Technology학부,

^o한국외국어대학교 Global Business & Technology학부

e-mail: jang110452@gmail.com*, {yelim3742**, park010302^o, dud1334**}@naver.com, yjmoon@hufs.ac.kr**

Prediction of the Number of Crimes according to Urban Environmental Factors in the Metropolitan Area

Ye-Won Jang*, Ye-Lim Kim**, Si-Hyeon Park^o, Jae-Young Lee**, Yoo-Jin Moon**

*Dept. of Italian Languages, Hankuk University of Foreign Studies,

**Division of Global Business&Technology, Hankuk University of Foreign Studies,

^oDivision of Global Business&Technology, Hankuk University of Foreign Studies

● 요약 ●

본 논문에서는 Scikit-learn 패키지의 LinearRegression 모델과 Keras 딥러닝 모델을 활용하여 수도권 도시 환경 요인에 따른 범죄 발생 건수를 예측 모델을 제안한다. 연구 방법으로 범죄 발생과 유의미한 관계가 있다고 파악되는 수도권의 각 자치구 별 데이터셋을 분석하여, CCTV, 파출소, 가로등의 수가 범죄 발생에 유의미한 영향을 끼치는 것을 확인하였다. 독립 변수들 간에 Scale을 줄이고자 정규화를 진행했고, 종속변수의 정규성 확보를 위해 로그변환을 취했다. 손실 함수는 회귀문제에서 사용되는 'relu' 함수를 사용했고 모델의 성능을 확인할 수 있는 지표로 MSE(Mean Squared Error)를 사용해 모델을 구성하였다. 본 논문에서 설계한 이 프로그램은 범죄 발생율이 높은 지역에 경찰 인력의 추가적 배치, 안전 시설 확충 등 실무적 조치를 취함에 있어 근거를 제공할 수 있을 것으로 사료된다.

키워드: 신경망(Neural Network), 회귀분석(Regression Analysis), 범죄 예측(Crimes Prediction)

I. Introduction

본 연구는 수도권을 대상으로 자치구별 범죄 발생 건수와 환경적 요인의 관계를 알아보고, 특정 지역에서의 범죄 발생 건수를 예측하고자 한다. 다수의 선행 연구 및 이론에서 범죄에 환경적 특성이 미치는 영향의 중요성이 강조되어 왔다. Cornish와 Clarke에 따르면, 지역의 환경적 특징은 인간의 의사결정에 영향을 미친다고 설명한다. 즉, 환경은 인간에게 범죄행위가 가능할 수 있도록 하는 기회를 제공한다. 또한 일상 활동이론에 따르면, 범죄는 동기가 부여된 범죄자가 보호자가 없는 적절한 피해를 시간과 공간적으로 동시에 만날 때 발생한다고 설명한다. 이처럼 범죄에 환경적 요인이 미치는 영향이 입증되어 왔으나 국내 범죄학 및 경찰학 관련 연구는 개인의 범죄성의 원인과 기회요인의 영향력 등에 대한 분석이 주를 이루어 범죄의 지리적 분포 및 환경적 요인과 관련해서는 공개된 자료가 한정적이며, 알려진 바가 거의 없다. 따라서 본 연구에서는 인공지능 알고리즘을 이용해 도시의 환경적 요인에 따른 범죄발생 건수를 모형화 하여 분석해보고자 한다[1-3].

II. Preliminaries

1. Related works

최근 TensorFlow를 이용한 범죄예측에 관한 다양한 연구가 진행되고 있다. 이전 연구는 약 10년 전 수집한 데이터를 활용한 프로그램이 주를 이룬다. 이에 비하여, 본 연구는 범죄 건수 예측을 위해 데이터를 2021년을 기준으로 수도권 도시의 데이터를 수집했으며, 자치구 별 범죄건수 예측을 하는 것에 중점을 두고 진행되었다. 또한 지역에 상관없이 독립변수와 종속변수로 설정한 데이터만 수집할 수 있다면 손쉽게 해당 지역의 범죄발생 건수를 예측할 수 있다. 데이터는 서울열린데이터광장, 공공데이터포털, 경찰청에서 제공하는 연도별, 지역별 범죄 발생 현황 자료를 가공하여 이용했다. 범죄 발생에 영향을 미칠 것이라고 판단한 'CCTV 수', '유형주점의 수', '파출소의 수' 등 11개의 독립변수와 종속변수인 '범죄발생 건수' 데이터를 수집했다.

2. Data sources

데이터는 서울열린데이터광장, 공공데이터포털, 경찰청에서 제공하는 연도별, 지역별 범죄 발생 현황 자료를 가공하여 이용했다. 범죄 발생에 영향을 미칠 것이라고 판단한 ‘CCTV 수’, ‘유흥주점의 수’, ‘과출소의 수’ 등 11개의 독립변수와 종속변수인 ‘범죄발생 건수’ 데이터를 수집했다.

III. The Proposed Scheme

1. Variables

범죄 발생 건수 예측의 정확도에 영향을 크게 미치는 변수를 판단하는 과정을 거쳤다. Fig. 1과 같다.

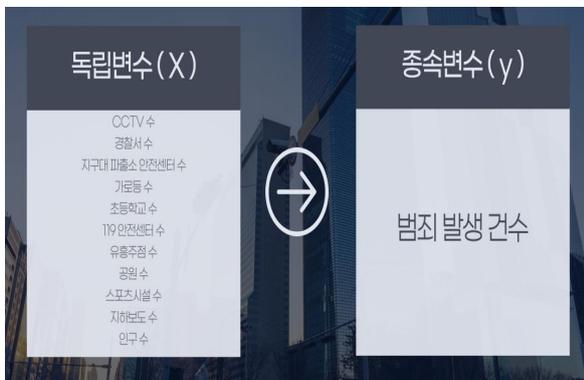


Fig. 1. Variable Data Set

2. Data normalization

독립변수들 간의 Scale 차이를 줄이고자 Fig. 2와 같이 정규화를 진행하여 종속변수(범죄발생 건수)를 로그 변환하여 정규성을 확보했다.

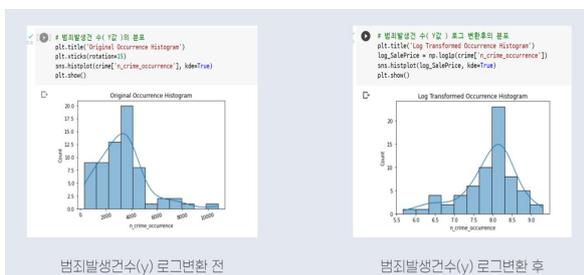


Fig. 2. Logarithm Transformation of Y Values

IV. Experiment

11개의 독립변수로 1개의 종속변수를 예측하는 다중회귀분석을 실시했다. Scikit-learn의 LinearRegression 모델을 사용했으며 하이퍼파라미터는 default값으로 진행했다. 회귀분석의 성능을 나타내는 결정계수값이 0.69로 확인된다. 즉, 해당 연구의 범죄예측 모형으로 실제 범죄발생건수의 약 69%정도 예측 가능하다는 것이다. 산점도를 통해 실제 값과 예측 값 사이에 유의미한 선형관계를 확인할 수 있었다.

V. Conclusions

본 연구를 통해 범죄 발생율이 높은 지역에 경찰 인력의 추가적 배치, 안전 시설 확충 등 실무적 조치를 취함에 있어 근거를 제공할 수 있을 것이다. 더불어 도시계획의 공간 구성이 야기하는 범죄발생과의 연계성을 구명함으로써 바람직한 도시계획 정립에 대안을 제시할 수 있을 것으로 보인다.

REFERENCES

- [1] Junmo Cho, “Big Data Analytics and Artificial Intelligence Starting with Python,” Infinity Books, 2020.
- [2] Jojo Moolayil, “Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python,” Apress, 2019.
- [3] Jun Wu, Jian Wang, Stephen Nicholas, Elizabeth Maitland, and Qiuyan Fan, “Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations”, Journal of Medical Internet Research, Vol. 22, No. 10: e21980, Oct. 2020. doi:10.2196/21980.