

「제5회 대구 빅데이터 분석 경진대회」 분석 결과 보고서

접수번호	※ 작성하지 않음
------	-----------

성명(팀명)	현풍담칼국수
분석과제명	신용평가모형을 활용한 대구지역 지역 소멸 등급 5등급 체계 구축 : 대구지역 활성화 방안

※ 20장 내외 자유 형식으로 작성, 목차는 필요시 변경 가능

I. 분석개요

□ 분석목적

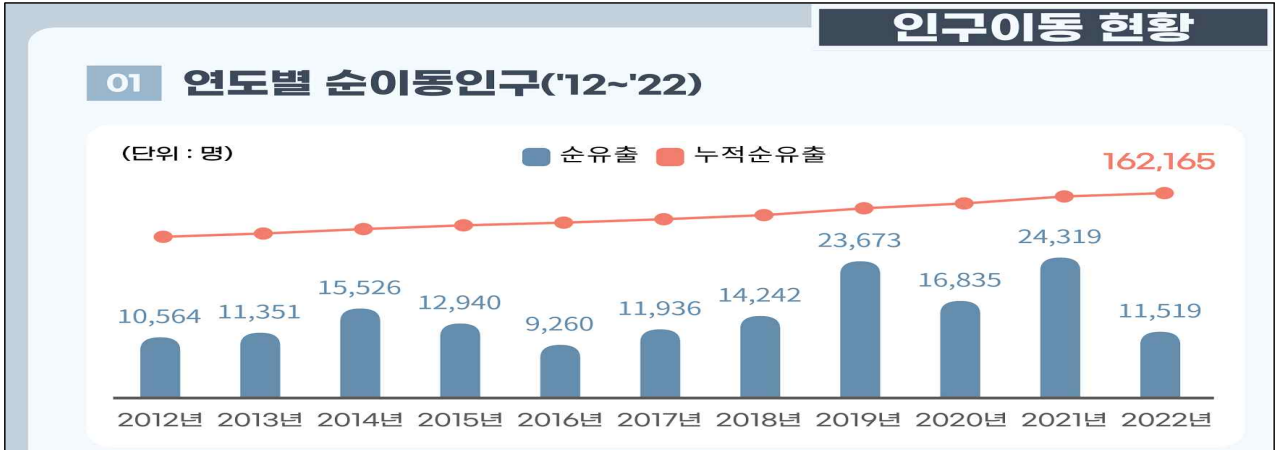
지방소멸의 위협이 점점 증가하는 현대사회에서, 대구광역시의 지역 활성화와 지역 소멸 방지는 중요한 과제로 대두되고 있습니다. 이러한 맥락에서 우리는 대구광역시의 읍면동 등급 체계를 구축하고 관련 요인을 분석하여 정책 가이드를 제공하고 지역 활성화 방안을 모색 하고자 합니다.

- ▷ (지방소멸 위험 지역 식별): 대구 지역의 읍면동 별로 신용평가모형을 활용하여 지방소멸 위험도를 분석하는 것입니다. 이를 통해 어떤 지역이 더 심각한 소멸 위험에 처해 있는지 정확하게 파악할 수 있습니다. 이는 각 지역마다 다른 문제점과 특성을 고려한 맞춤형 정책 제안에 필요한 기본적인 정보를 제공합니다.
- ▷ (대구광역시 정책 방향 제공): 분석 결과를 바탕으로 대구광역시의 각 읍면동 별로 적합한 정책 가이드라인을 제시하는 것입니다. 이는 현재와 미래의 지방소멸 문제 해결에 필요한 구체적인 방안 마련에 크게 기여합니다.
- ▷ (신용 평가 모형 활용 목적): 금융기관에서 주로 사용되는 신용 평가 모델을 새롭게 해석하여, 사회문제인 '지방소멸' 문제에 적용함으로써 차별화된 접근법을 시도하는 것입니다. 이는 기존 경제 지표나 인구 통계 등 일반적인 예측 방법이 아닌, 과학적 데이터 분석 및 예측 모델링으로 보다 정확하고 섬세한 예측 결과를 도출할 수 있습니다.
- ▷ (Tableau 대시보드 구축 목적): 복잡한 신용 평가 모델링 결과를 직관적으로 이해할 수 있는 Tableau 대시보드를 구축하여, 관련 정보를 한눈에 확인하고 비교분석할 수 있도록 하는 것입니다. 이를 통해 정책 결정자나 관련 기관, 그리고 일반 시민들도 쉽게 데이터를 해석하고 활용할 수 있습니다.

□ 배경

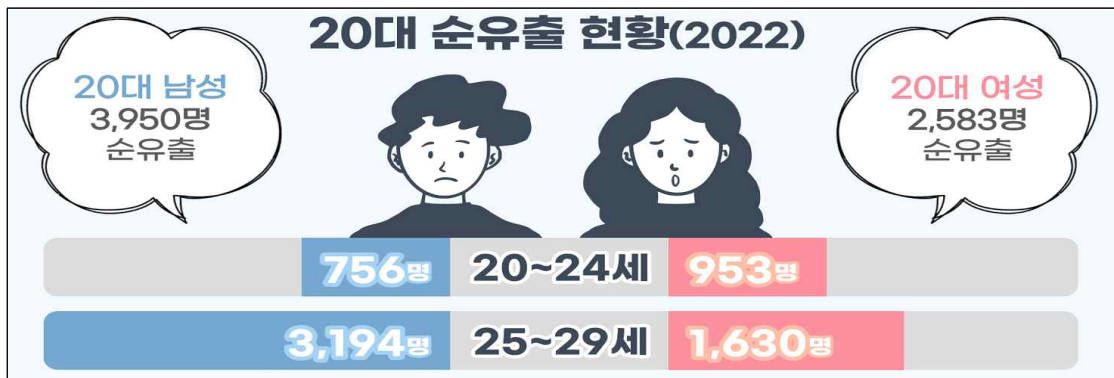
○ 지속적인 지역 인구감소

- ▷ 지역사회의 안정적 발전과 지속 가능성은 그 지역의 인구 동태에 크게 의존합니다. 그러나 최근 대구광역시에는 심각한 인구 감소 문제에 직면해 있습니다.



○ 청년인구의 수도권 쏠림 현상

- ▷ 2022년 기준으로 대구에서 빠져나간 인구 순유출이 1만 1519명에 이르렀으며, 이 중 20대 청년층의 유출이 가장 많았습니다. 대기업 공장 및 본사, 연구소 등이 부족하고 중소기업 비중이 높은 산업 구조 때문에 주요 대학교(경북대, 계명대, 영남대 등)의 졸업생들이 다른 지역으로 유출되는 현상도 지속되고 있습니다.



□ 필요성

○ 지방 소멸 위험 등급화

- ▷ **(정책 방향성 제시):** 우리는 어떤 읍면동이 더 심각한 지방소멸 위험에 처해 있는지 식별하고 해당 지역에서 어떤 정책을 우선적으로 시행해야 할지 방향성을 제시할 수 있습니다.
- ▷ **(정책 목표 설정):** 각 분야별 문제점을 구체적으로 파악하여 그에 따른 정책

목표를 설정할 수 있습니다. 이를 통해 각 지역의 특성에 맞는 방안을 마련하고 실행함으로써 지방 소멸을 예방할 수 있습니다.

- ▷ **(정책 성과 평가):** 정부 및 관련 기관은 등급화된 결과를 바탕으로 정책 성과를 모니터링하고 평가할 수 있습니다. 특정 읍면동의 등급 변화를 통해 정책 효과를 확인하고 필요 시 조치 조정을 할 수 있습니다.

따라서, 본 프로젝트는 지방소멸 위험도 등급화를 통해 실질적인 정책 방향성을 제시하고, 이를 바탕으로 정책 목표를 설정하며, 실행된 정책의 성과를 평가하는데 필요한 중요한 도구로 작용할 것입니다.

□ 분석요약

○ 활용데이터 :

(인구부분) 주민등록인구 및 세대 현황, 고령 인구현황, 20~39세 여성인구 (경제부분) 현대카드 나이대별 매출 데이터, 총 종사자수 (생활서비스) 은행 종사자수, 부동산업 종사자수 (여가생활) 생활체육시설 종사자수, 노래방 종사자수 (음식) 카페 종사자수, 패스트푸드 종사자수 (교육) 기술직업훈련 종사자수, 어린이보육원 종사자수, 초등학교 종사자수, 중학교 종사자수, 고등학교 종사자수, 어학원 종사자수 (교통) 이륜차, PM 사고건수, 교통량 (공공) 경찰 종사자수, 행정기관 종사자수

○ 분석도구 : Python, QGIS, Tableau 등

○ 분석기법 :

- (1) 신용평가모형 개발(scorecardpy)
- (2) 로지스틱 회귀모형
- (3) IV(Information Value) Selection
- (4) WOE Binning, fine classing(계급 세분화)
- (5) coarse classing(성김화)
- (6) k-s통계량 Selection
- (7) 독립 표본 t-검정(Independent Samples t-test) Selection
- (8) 스피어만 상관 계수(Spearman correlation coefficient) 다중공선성 확인
- (9) 라쏘(LASSO) Selection
- (10) AUROC, K-S, PSI 모형 성능 및 안정성 평가
- (11) score에 대한 등급화 - 위험률에 대한 역전 현상 확인

○ 분석결과

- 본 연구는 신용평가 모형의 원리를 새롭게 해석하여 지방소멸 위험도를 정밀하게 평가하는 독창적인 방법을 제시하였습니다.
- 이 모형은 소멸 위험도를 정량적으로 나타내는 동시에, 각 등급별로 정책 가이드라인을 제공함으로써 현재와 미래의 지방소멸 문제 해결에 대한 구

체적인 방안 마련에 크게 기여할 수 있습니다.

- 역전 현상이 보이지 않은 지역 소멸 위험 등급 5등급 체계를 구축했습니다.

○ 활용방안 :

- **(지방소멸 위험도 관리 도구):** 본 프로젝트에서는 지방소멸 위험도를 정량화하여 신용 점수처럼 등급으로 분류하였습니다. 이렇게 구축된 모델은 지역 소멸 위험을 시각적이고 직관적인 방식으로 이해하는 데 도움이 됩니다. 각 지역의 소멸 위험도를 정확하게 산출함으로써, 지역 소멸 문제에 대한 인식을 높이고, 이에 따른 예방 및 대응 전략을 수립하는데 중요한 참고 자료가 될 수 있습니다.
- **(정부 및 관련 기관의 정책 마련 참고 자료):** 본 프로젝트에서 제공하는 각 등급별 정책 가이드라인은 현재와 미래의 지방소멸 문제 해결에 필요한 구체적인 방안 마련에 크게 기여합니다. 특정 지역의 소멸 위험도 등급에 따라 적합한 정책 가이드라인을 제시함으로써, 해당 지역들의 장기적 발전과 활성화 전략 수립에 있어서 중요한 근거 자료를 제공합니다.
- **(지역 소멸 위험 등급 Tableau 대시보드 구축):** 복잡한 신용평가모형에 대한 정보를 직관적이고 이해하기 쉬운 형태로 변환하기 위해 Tableau 대시보드를 구축하였습니다. 아래와 같이 개별 읍면동의 소멸 위험도 점수와 등급, 그리고 지역 소멸 위험 지수 추세까지 한 눈에 확인할 수 있는 대시보드를 구축하였습니다.



□ 독창성 및 차별성

- ▷ **(새로운 접근법의 도입):** 해당 공모전 프로젝트의 독창성과 차별성은 보통 금융 기관에서 주로 활용되는 신용평가모형 개발 방법을 완전히 새로운 시각과 분야에 적용한 점에 있습니다. 우리는 지방소멸 위험을 평가하기 위해 신용평가모형의 원리를 활용하여 지방소멸 위험지역과 비위험지역을 분류하였습니다. 이는 기존에는 경제 지표나 인구 통계 등을 토대로 한 예측이 아닌, 신용평가 모델을 통한 예측으로써 새로운 시각을 제시하는 것입니다.

- ▷ (‘Scorecardpy’를 활용한 지방소멸 위험도 점수화 및 등급 분류: 정량적 평가 체계의 구축): 이 프로젝트의 핵심은 파이썬 라이브러리인 ‘scorecardpy’를 사용하여 신용평가모형을 구축한 것입니다. 우리는 예측 확률을 통해 각 지역의 소멸 위험도를 점수화하고, 이 점수를 기준으로 안정적인 등급 분류 시스템을 만들었습니다. 이러한 방식은 지방소멸 위험도를 정량적 수치로 나타내는 동시에, 각각의 지역 경제 상태에 대한 예측력 있는 평가 체계를 제공합니다.
- ▷ (등급별 정책 가이드라인 제공: 지방소멸 문제 해결을 위한 구체적 방안 마련과 장기적 발전 전략): 더욱 중요한 것은, 본 프로젝트가 단지 소멸 위험이 높은 지역과 그렇지 않은 지역을 분류하는 것에서 그치지 않고, 각 등급별 정책 가이드라인까지 제공한다는 점입니다. 이는 현재와 미래의 지방소멸 문제 해결에 대한 구체적인 방안 마련에 크게 기여할 수 있으며, 궁극적으로 해당 지역들의 장기적 발전과 활성화 전략 수립에 사용될 수 있습니다.

따라서 본 프로젝트에서 개발된 모형은 기존 신용 평가 방법론을 새롭게 해석하여 적용함으로써 차별화되어 있으며, 과학적인 데이터 분석 및 예측 모델링과 함께 사회 문제 해결에 필요한 실질적인 가이드라인을 제공함으로써 독창성을 발휘하고 있습니다. 이러한 접근법은 금융기관뿐만 아니라 정부 및 관련 기관에서도 지역의 소멸 위험도를 평가하고 대응하기 위한 중요한 참고 자료로 활용될 것입니다.

II. 분석방법

□ 활용데이터

유형	데이터명	기간	상세정보	출처
인구부문	연령별인구현황	2019.01~ 2022.06	20~39세 여성인구수	 행정안전부
	주민등록인구기타현황 (고령 인구현황)	2019.01~ 2022.06	월별 전체 고령인구수	 행정안전부
	주민등록 인구 및 세대 현황 (월간)	2019.01~ 2022.06	세대수, 세대당 인구, 남자 인구수, 여자 인구수, 남녀 비율, 거주자 인구수	 행정안전부
소매업	백화점/중대형마트 종사자수	2019.01~ 2022.06	백화점/중대형마트 종사자수	 SGIS ^{plus} 통계지리정보서비스
	편의점 종사자수	2019.01~ 2022.06	노래방 종사자수	 SGIS ^{plus} 통계지리정보서비스
경제부문	현대카드 데이터	2019.01~ 2022.08	나이대별(10대, 20대., 30대, 40대, 50대., 60대), 성별(남,여) 매출액	 DAEGU BIG DATA 대구 빅데이터 활용센터
	종사자수	2019.01~ 2022.06	종사자수	 SGIS ^{plus} 통계지리정보서비스
생활 서비스	은행 종사자수	2019.01~ 2022.06	은행종사자수	 SGIS ^{plus} 통계지리정보서비스
	부동산업 종사자수	2019.01~ 2022.06	부동산 종사자수	 SGIS ^{plus} 통계지리정보서비스
여가생활	생활체육시설 종사자수	2019.01~ 2022.06	생활체육시설 종사자수	 SGIS ^{plus} 통계지리정보서비스
	노래방 종사자수	2019.01~ 2022.06	노래방 종사자수	 SGIS ^{plus} 통계지리정보서비스
음식	카페 종사자수	2019.01~ 2022.06	카페 종사자수	 SGIS ^{plus} 통계지리정보서비스
	패스트푸드 종사자수	2019.01~ 2022.06	패스트푸드 종사자수	 SGIS ^{plus} 통계지리정보서비스

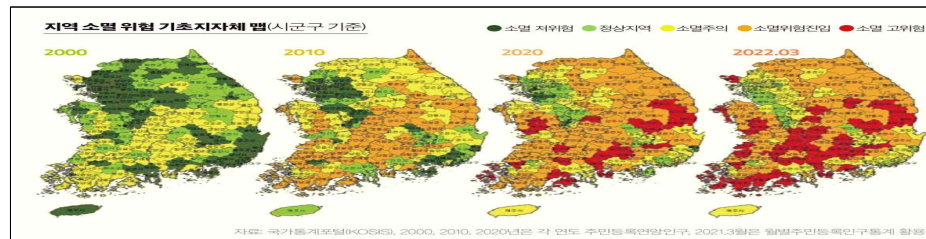
교육	기술 직업 훈련 종사자수	2019.01~ 2022.06	기술 직업 훈련 종사자수	SGIS ^{plus} 통계지리정보서비스
	어린이보육업 종사자수	2019.01~ 2022.06	어린이보육업 종사자수	SGIS ^{plus} 통계지리정보서비스
	초등학교 종사자수	2019.01~ 2022.06	초등학교 종사자수	SGIS ^{plus} 통계지리정보서비스
	중학교 종사자수	2019.01~ 2022.06	중학교 종사자수	SGIS ^{plus} 통계지리정보서비스
	고등학교 종사자수	2019.01~ 2022.06	고등학교 종사자수	SGIS ^{plus} 통계지리정보서비스
	여학원 종사자수	2019.01~ 2022.06	여학원 종사자수	SGIS ^{plus} 통계지리정보서비스
공공	경찰 종사자수	2019.01~ 2022.06	경찰 종사자수	SGIS ^{plus} 통계지리정보서비스
	행정기관 종사자수	2019.01~ 2022.06	행정기관 종사자수	SGIS ^{plus} 통계지리정보서비스
교통	이륜차, PM 사고건수	2019.01~ 2022.06	이륜차 사고건수 + Personal Mobility 사고건수	TAAIS 교통사고분석시스템 Traffic Accident Analysis System
	도로 정보 및 교통량	2017.01 ~ 2018.12	승용차 교통량 + 버스 교통량 + 트럭 교통량	DAEGU BIG DATA 대구 빅데이터 활용센터
건축	아파트 실거래 매매 횟수	2019.01~ 2022.06	아파트 매매 거래 횟수	국토교통부
	지가변동률	2019.01~ 2022.06	지가변동률	KOSIS

□ 분석방법

○ 데이터 수집 및 구축

1) 종속 변수 정의

소멸위험지수란 만 20~39세 여성 인구를 만 65세 이상 인구로 나누어 낸 지수로 일본의 사회학자 마스다 히로야가 처음 개념을 정의 하였습니다. 이 지수가 1.5 이상이면 소멸 위험이 매우 낮은 소멸 저위험 지역, 1.0~1.5인 경우 보통, 0.5~1.0인 경우 주의, 0.2~0.5는 소멸 위험, 0.2 미만은 소멸 고위험 지역으로 분류됩니다. 소멸위험지수가 0.5 미만이면 소멸위험지역이라고 정의합니다. 우리는 각 동별로 만 65세 인구와 만 20~29 여성인구를 추출하여 지방소멸지수를 계산하였습니다. 그리고 ‘위험여부’ 라는 칼럼을 만들어 지방소멸 지수가 0.5 미만인 지역들은 1이라는 값을 부여 하였고 0.5 이상인 지역들은 0이라는 값을 부여하였습니다.



2) 독립 변수 선정 기준

대구광역시의 읍면동 등급 체계를 구축하고 관련 요인을 분석하여 정책 가이드를 제공하고 지

역 활성화 방안을 모색하는 것이 본 분석의 목적이기 때문에 어느 부분이 부족하고 활성화 해야 하는지에 대해 판단을 하기위해 인구부문, 경제부문, 소매업, 생활서비스, 여가생활, 음식, 공공, 교통, 건축으로 분류를 하고 독립변수들을 선정하였습니다. 인구부문은 지역 내 인구 구조 및 동향을 분석하고 인구 변화에 따른 사회 및 경제 영향을 예측하기 위해 1인가구수, 2인가구수, 3,4인가구수, 5,6인가구수, 7인가구수, 상주인구, 세대수, 세대당 인구수, 남자인구수, 여자인구수, 남녀비율의 독립변수들을 선정 하였습니다. 경제부문은 총 산업 분야의 종사자 수와 그 지역의 매출을 토대로 하여 경제 활성화와 기술 발전을 예측하기 위해 나이대별 매출액과 종사자수를 독립변수로 선정하였습니다. 소매업은 상업 시설 활성화 정도를 분석하여 소상공인의 역할을 파악하기 위해 백화점/중대형마트 종사자수와 편의점 종사자수를 독립변수로 선정하였습니다. 생활 서비스는 생활 서비스를 분석하여 지역 내 생활 편의성과 품질을 예측하기 위해 은행 종사자수와 부동산 종사자수를 독립변수로 선정하였습니다. 여가생활은 문화 시설 발전과 주민의 여가 생활 풍요를 예측하기 위해 생활체육시설 종사자수와 노래방 종사자수를 독립변수로 선정하였습니다. 공공은 행정, 안전, 공공 서비스의 수준을 파악하여 주민의 안전과 행정 서비스 품질을 예측하기 위해 경찰 종사자수와 행정기관 종사자수를 독립변수로 선정하였습니다. 음식은 지역 내 음식 서비스의 활성화 정도를 예측하기 위해 카페 종사자수와 패스트푸드종사자수를 독립변수로 선정하였습니다. 교육은 교육 분야의 활성화 정도와 교육 인프라를 파악하며, 지역 내 교육 환경과 교육 수준의 발전을 예측하기 위해 어린이 보육업 종사자수, 초등학교 종사자수, 중학교 종사자수, 고등학교 종사자수, 어학원 종사자수를 독립변수들로 선정하였습니다. 교육 분야의 활성화 정도와 교육 인프라를 파악하며, 지역 내 교육 환경과 교육 수준의 발전을 예측하기 위해 교통량과 이륜차, pm사고건수를 독립변수로 선정하였습니다. 건축은 부동산 시장과 주거 환경을 파악하며, 주택 시장 동향과 주거 환경 개선을 예측하기 위해 아파트 실거래 매매 횟수와 지가변동률을 독립변수로 선정하였습니다.

3) 데이터 병합 및 데이터 시계열

각 분야별 데이터들을 하나의 데이터 프레임으로 합치기 위해 하위동들을 최상위 행정동으로 변환하여 행정동,년도,분기를 기준으로 sum을 하여 독립변수들을 계산했습니다. 혁신동과 유천동은 각 2020년 10월 , 2021년 11월에 생긴 동으로 현대카드 데이터에 안심동과 진천동으로 카운트가 되어있어 혁신동과 유천동은 안심동, 진천동으로 변환하여 계산하였습니다. 지방소멸위험지수, 지가변동률, 세대당인구수, 남녀비율은 sum한 값이 아니라 비율이기 때문에 하위동에서 최상위 행정동으로 바뀐 수 만큼 나눠주어 평균으로 만들어 주었습니다.

○ 신용평가모형 개발 과정 - scorecardpy in python

1) scorecardpy이란 ?

- scorecardpy는 파이썬에서 사용할 수 있는 신용 점수카드 모델을 개발하는 데 사용되는 라이브러리입니다. 이를 통해 우리는 데이터의 정보 가치(Information Value,

IV)를 측정하고, 최적의 분할을 찾아내며, WOE 변환과 점수카드 생성 등을 진행하여 신용평가 모형을 모델링할 수 있습니다.

2) IV(Information Value)을 이용한 1차 변수 Selection

<제거된 변수: '1인가구', '2인가구', '3,4인가구', '남자인구', '상주인구', '세대수', '여자인구' >

- 정보 가치(IV)는 예측 변수가 얼마나 유의한지 판단하는 지표입니다. IV가 높은 변수일수록 타겟 변수를 예측하는데 도움이 됩니다. 따라서 우리는 IV를 계산하여 가장 중요한 변수들을 선택합니다.
- 본 프로젝트에서는 독립변수의 수가 많고, 변수 선택을 총 4회에 걸쳐 진행하기 때문에, 미리 높은 기준으로 변수를 제거하는 작업을 실시하였습니다. 구체적으로는 정보 가치(IV)가 0.3 미만인 변수들은 제거하였습니다. 이렇게 하면 비교적 중요도가 낮은 변수들을 사전에 제외함으로써 모델의 복잡성을 줄이고 계산 효율성을 향상시킬 수 있습니다.
- 그러나 이 과정에서 한 가지 예외 사항이 있었습니다. 바로 '현대 카드 데이터의 나이대별 매출액' 변수입니다. 이 변수는 중요도가 높다고 판단하여 1차 변수 선택에서 제외하였습니다. 즉, 이 변수에 대해서는 패널티를 완화하여 2차 변수 선택에서부터 변수 제거를 고려했습니다.

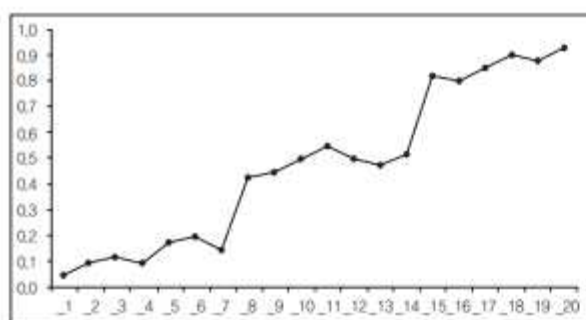
$$IV = \sum [(\%Good - \%Bad)] \times \ln \left(\frac{\%Good}{\%Bad} \right)$$

$$= \sum_{j=1}^l \left[\frac{n_{1j}}{n_1} - \frac{n_{0j}}{n_0} \right] Woe_j$$

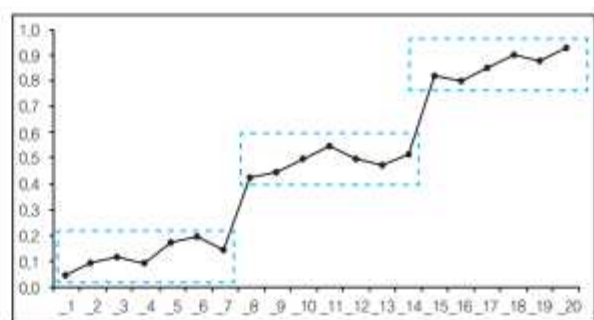
$IV \leq 0.02$: 예측불능
 $0.02 < IV \leq 0.1$: 약한 예측력
 $0.1 < IV \leq 0.3$: 중간 예측력
 $0.3 < IV$: 강한 예측력

<IV(Information Value) 계산식>

3) WOE Binning, fine classing(계급 세분화)



(a) 계급세분화 - 최대 20개 구간



(b) 성김화 - 재범주화

신용평가 모형에 있어 필수적인 계급 세분화와 성김화 작업은 필수적입니다.

1) 계급세분화(Fine Classing): 계급세분화는 원래의 독립변수 값을 종속변수인 '불량'과의 관계를 분석하여 불량률이 유사한 범주를 하나로 묶어 분석에 사용하는 방법입니다. 이 과정은 크게 두 가지 목적이 있습니다.

- 변수 간 관계 파악: 세분화된 계급을 통해 독립변수와 종속변수 간의 관계를 보다 명확하게 이해할 수 있습니다. 즉, 어떤 범주가 '불량'에 더 큰 영향을 주

는지 파악할 수 있습니다.

- 모델 예측력 향상: 세분화된 계급은 모델의 예측력을 향상시키는데 중요한 역할을 합니다. 이는 각 그룹 내에서 데이터 패턴이 보다 일관되기 때문입니다.

2) 성김화(Coarse Classing): 성김화는 세분화된 클래스들을 합치거나 재구성하여 보다 일반적인 범주로 만드는 과정입니다. 이 과정은 다음과 같은 목적으로 진행됩니다.

- 모델 안정성 및 해석력 향상: 너무 많은 클래스가 있는 경우, 모델이 데이터에 과적합될 위험이 있고, 결과 해석이 복잡해질 수 있습니다. 따라서 적절한 수준에서 클래스를 재구성함으로써 모델의 안정성과 해석력을 향상시킬 수 있습니다.
- 일반화(generalization)능력 강화: 너무 상세하게 나뉜 범주들을 합치면서 신규 데이터에 대한 모델의 일반화 능력도 개선합니다.

3) WOE 기반 계급 세분화:

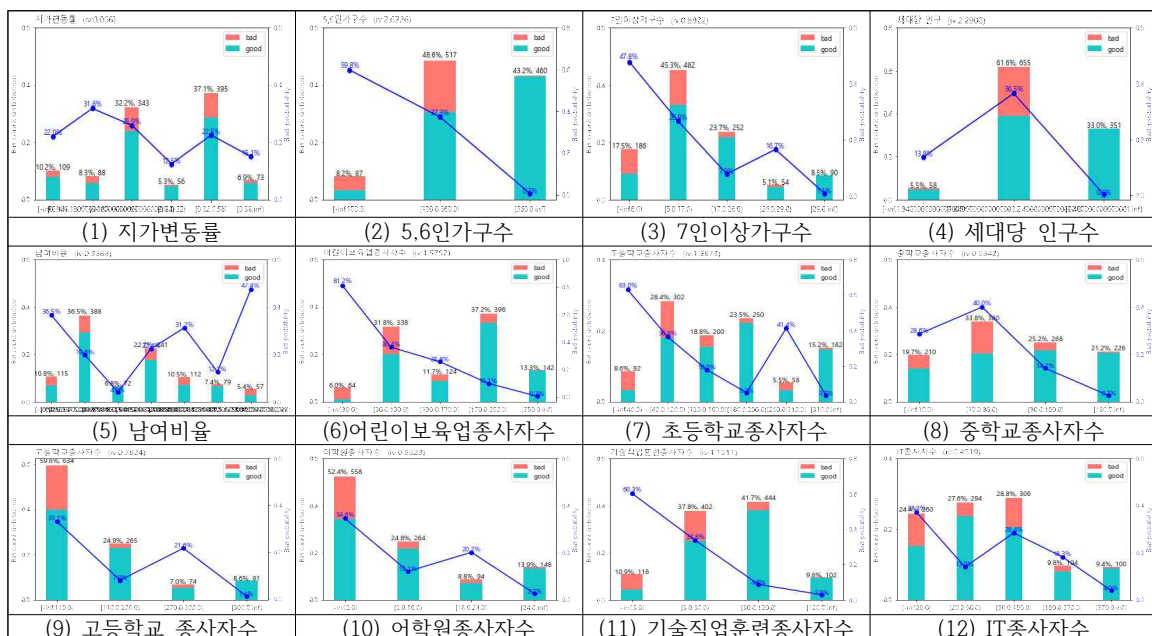
- 해당 프로젝트에서는 , Weight of Evidence(WOE)를 기준으로 독립변수 값을 계급화합니다. WOE는 종속변수인 '불량'과의 관계를 분석하여 불량률이 유사한 범주를 하나의 범주로 묶은 후, 이를 다양한 구간으로 세분화하는 방법입니다.

$$WoE_j = \ln \left(\frac{\%Good}{\%Bad} \right)$$

$$= \ln \left[\frac{\frac{n_{1j}}{n_1}}{\frac{n_{0j}}{n_0}} \right] = \ln \frac{n_{1j}}{n_{0j}} - \ln \left[\frac{n_1}{n_0} \right]$$

<woe 계산식>

<WOE Binning, fine classing - 계급세분화 결과>





4) k-s통계량, 독립 표본 t-검정(Independent Samples t-test) 2차 변수 Selection

<제거된 변수: '지가변동률' >

- 위에서 진행한 계급 세분화에 대한 구간에 대해서 ks통계 검정과 독립 표본 t검정을 통해서 위험 여부를 판별할 수 있는 유의미한 변수인지 통계 검정을 수행합니다.

- KS 통계량은 두 분포 간의 차이를 측정하는 방법으로, '불량'과 '비불량' 그룹 간에 어느 변수가 가장 큰 차이를 보이는지 판단하는 데 사용됩니다.
- 독립 표본 t-검정은 두 집단 간 평균의 차이가 유의미한지 검증하는 방법으로, 여기서는 '불량'과 '비불량' 그룹에서 해당 변수의 평균 값이 유의미하게 다른지 확인하는 데 사용됩니다.
- 따라서 WOE 기반 계급세분화와 KS 통계량 및 t-검정을 활용하여 위험 여부 판단에 중요한 역할을 하는 변수들을 선택하였습니다. 이러한 과정은 신용평가모형에서 중요한 역할을 하는 독립변수 선택 단계입니다.

< k-s통계량, 독립 표본 t-검정 결과 >

Variabal	KS_Statistic	KS_PValue	T_Statistic	T_PValue
중학교종사자수	0.37	< 0.0001	-9.19	< 0.0001
매매횟수	0.44	< 0.0001	-10.93	< 0.0001
IT종사자수	0.26	< 0.0001	-6.91	< 0.0001
어학원종사자수	0.32	< 0.0001	-8.05	< 0.0001
남여비율	0.17	0.00037	-5.73	< 0.0001
세대당 인구	0.42	< 0.0001	-10.94	< 0.0001
패스트푸드종사자수	0.39	< 0.0001	-13.34	< 0.0001
노래방종사자수	0.37	< 0.0001	-15.98	< 0.0001
종사자수	0.49	< 0.0001	-14.99	< 0.0001
백화점,중대형마트종사자수	0.13	0.01726	-4.69	< 0.0001
어린이보육업종사자수	0.44	< 0.0001	-12.35	< 0.0001
.				
부동산종사자수	0.55	< 0.0001	-17.65	< 0.0001
병원종사자수	0.36	< 0.0001	-9.58	< 0.0001
행정기고나종사자수	0.23	< 0.0001	-6.51	< 0.0001
생활체육시설종사자수	0.33	< 0.0001	-11.82	< 0.0001
기술직업훈련종사자수	0.43	< 0.0001	-11.93	< 0.0001
고등학교종사자수	0.34	< 0.0001	-8.34	< 0.0001
카페종사자수	0.36	< 0.0001	-12.64	< 0.0001
7인이상가구수	0.30	< 0.0001	-8.54	< 0.0001
은행종사자수	0.48	< 0.0001	-13.94	< 0.0001
금융업종사자수	0.41	< 0.0001	-12.79	< 0.0001
5.6인가구수	0.54	< 0.0001	-14.75	< 0.0001
지가변동률	0.05	0.7804	-24.43	0.015

* ks_T_statistics.csv 참조

5) coarse classing(성김화)

- 위에서 설명한 성김화를 진행한 결과는 아래와 같습니다.
- 성김화(Coarse Classing):

성김화는 계급세분화를 통해 생성된 세부 범주들을 재구성하여 보다 일반적인 범주로 만드는 과정입니다. 이는 모델의 안정성을 높이고, 해석력을 강화하는데 중요합니다.

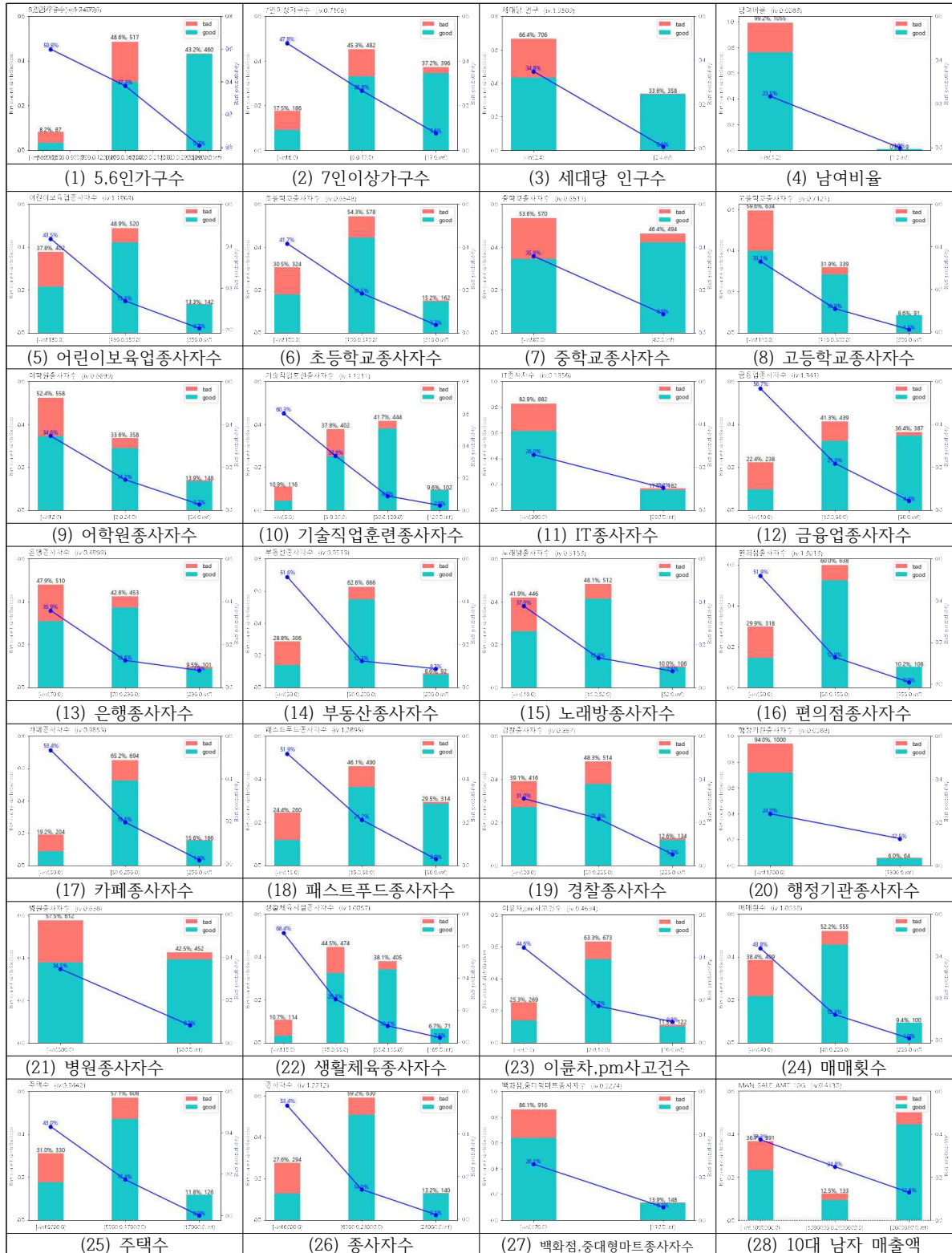
본 프로젝트에서는 모든 변수에 대해 불량률에 역전 현상이 없도록 구간을 재범주화하였습니다.

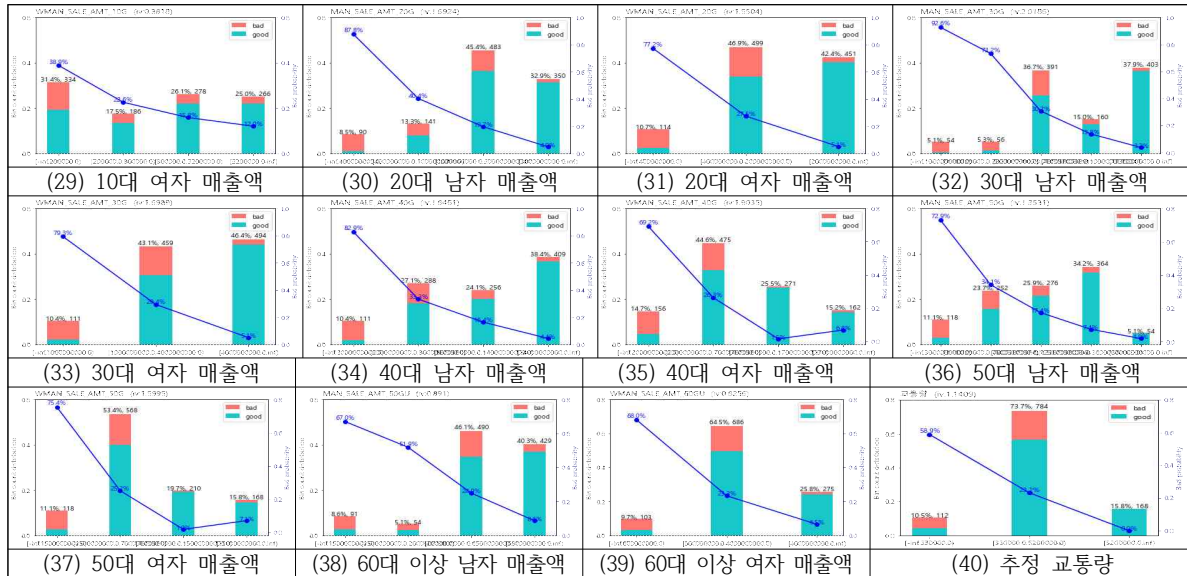
- 역전 현상이 없다는 것은 불량률이 구간이 증가함에 따라 불량률이 증가하거나 감

소하는 단조로운 형태를 띠는 것을 의미합니다.

- 모든 변수에 대한 수동적인 구간화를 통해 역전현상을 제거하여, 성검화를 진행해 주었습니다.

<coarse classing - 성검화 결과>





6) 스피어만 상관 계수(Spearman correlation coefficient) 다중공선성 확인 3차 변수 Selection

<제거된 변수: '어린이보육업종사자수_woe', '편의점종사자수_woe', '주택수_woe',

"MAN_SALE_AMT_30G_woe", "MAN_SALE_AMT_40G_woe",
 "MAN_SALE_AMT_60GU_woe", "MAN_SALE_AMT_50G_woe",
 "WMAN_SALE_AMT_40G_woe", "WMAN_SALE_AMT_50G_woe">

- 본 연구에서는 각 독립변수들의 관계를 분석하기 위해 스피어만 상관계수를 활용하였습니다. 스피어만 상관계수는 순서형 변수(ordinal variable)인 계급 구간과 같은 데이터에 적합한 비모수적 방법입니다.
- 이를 통해, 각 변수들이 서로 높은 상관관계를 가지고 있는지 확인하였으며, 이 중 중요하다고 판단되는 변수만을 최종적으로 선별하여 다중공선성 문제를 방지하였습니다.
- 다중공선성(Multicollinearity)은 두 개 이상의 독립변수가 서로 높은 상관관계를 가질 때 발생하는 문제입니다. 이런 현상이 발생할 경우, 다음과 같은 문제가 있을 수 있습니다.

- 1) 모델의 안정성 저하: 공선성이 높은 변수들이 모델에 포함될 경우, 회귀 계수 추정치의 분산이 커져 모델의 안정성을 저해합니다.
- 2) 해석력 감소: 다중공선성 때문에 어떤 독립변수가 종속변수에 미치는 영향력을 정확히 파악하기 어렵게 만듭니다.
- 3) 과적합 위험 증가: 다중공선성 문제는 모델이 과도하게 복잡해져 실제 데이터에서 벗어난 예측 결과를 초래할 수 있습니다.

따라서, 본 연구에서는 서로 높은 상관관계를 가진 변수들 중 중요도가 높은 변수만을 선택하여 사용함으로써, 심화된 분석 단계에서 생길 수 있는 다중공선성 문제를 사전에 방지하였습니다. 이러한 절차는 모델의 예측력 향상 및 해석력 강화에 크게 기여합니다.

8) AUC , ks통계량 모형 평가

- 본 연구에서 구축한 로지스틱 회귀 모형의 성능을 평가하기 위해 KS 통계량과 AUC를 사용하였습니다.

<KS 통계량>:

- 학습 데이터의 KS 통계량은 0.7824로, 검증 데이터의 KS 통계량은 0.7942로 나타났습니다.

:KS 통계량은 '불량'과 '비불량' 그룹 간 분포 차이를 측정하는 지표로, 값이 클수록 그 차이가 크다는 것을 의미합니다.이 결과는 모형이 두 그룹을 잘 구분하고 있음을 보여줍니다.

<AUC(Area Under the ROC Curve)>:

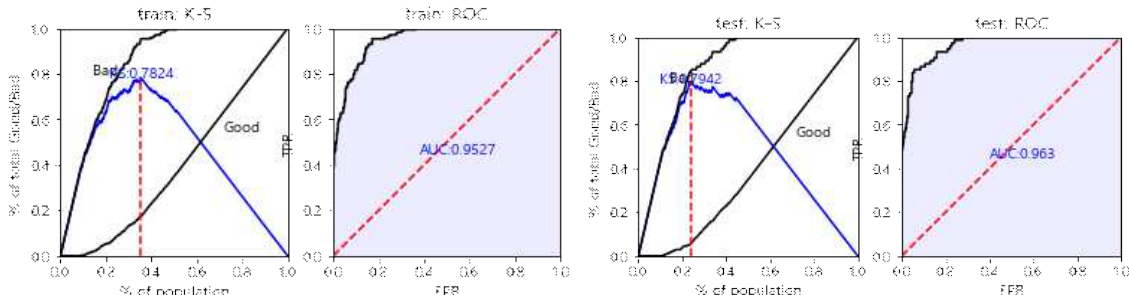
- 학습 데이터의 AUC는 0.9527로, 검증 데이터의 AUC는 0.963으로 나타났습니다.

: AUC 값은 모형이 클래스를 올바르게 예측하는 능력을 측정하는 지표입니다. 이 결과는 모형이 높은 예측 성능을 가지고 있음을 보여줍니다.

- 따라서 본 연구에서 구축한 모형은 매우 우수한 성능을 보여주며, 과적합(Overfitting) 혹은 과소적합(Underfitting) 문제도 발생하지 않았음을 확인할 수 있습니다.
- 추가 평가 및 주의사항

: 그러나 이러한 결과에도 불구하고 다음과 같은 사항들에 주의해야 합니다:

- 새로운 데이터에 대한 일반화: 현재까지의 평가 결과는 기존 학습 및 검증 데이터에 대한 것입니다. 따라서 신규 데이터에서도 동일한 성능이 유지될 수 있는지 계속해서 확인해야 합니다.
- 모형 업데이트와 유지보수: 시간 경과와 함께 환경 변화 등으로 인해 모형 성능 저하 가능성이 있으므로, 정기적인 모형 업데이트와 유지보수 작업 필요성에 대해 인식해야 합니다.
- 이러한 점들을 고려하면서 본 모형의 활용과 향후 개선 방향을 계획하면, 더욱 효과적인 신용 평가를 수행할 수 있을 것입니다.

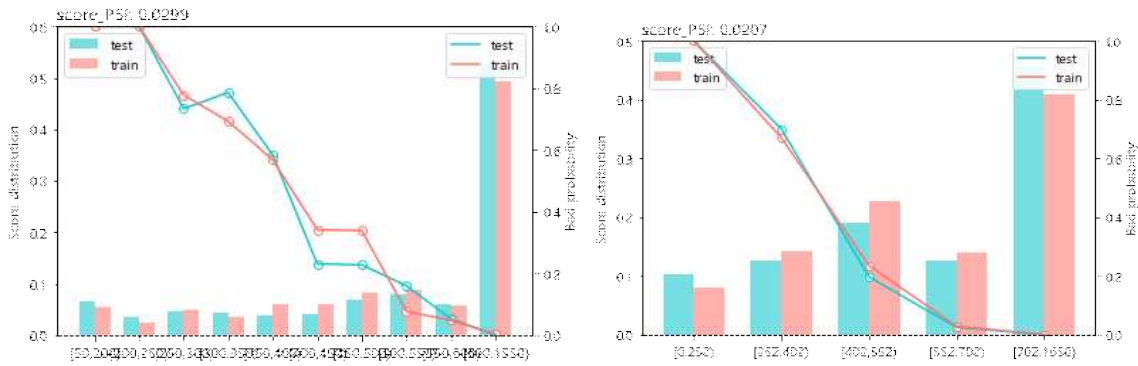


9) score에 대한 등급화 - 위험률에 대한 역전 현상 확인

- 본 연구에서는 로지스틱 회귀 모형을 통해 개발된 스코어 카드를 산출하였습니다. 이 소멸 위험지역에 대한 스코어 카드는 대구 광역시의 읍면동에 적용되었습니다.

- 스코어 카드의 등급 체계는 다음과 같이 구축하였습니다.

왼쪽 score 등급이 역전 현상을 고려하지 않고 구축한 등급화 시각화이며, 오른쪽 score 등급이 역전 현상을 고려하여 소멸 위험 등급 5등급 체계를 구축한 모습입니다.



- 등급 체계 구축 설명:

스코어 카드 점수를 기준으로 고객을 5개 등급으로 분류하였습니다.

각 등급은 고객의 신용 위험도를 나타내며, 1등급이 가장 낮은 위험, 5등급이 가장 높은 위험을 의미합니다.

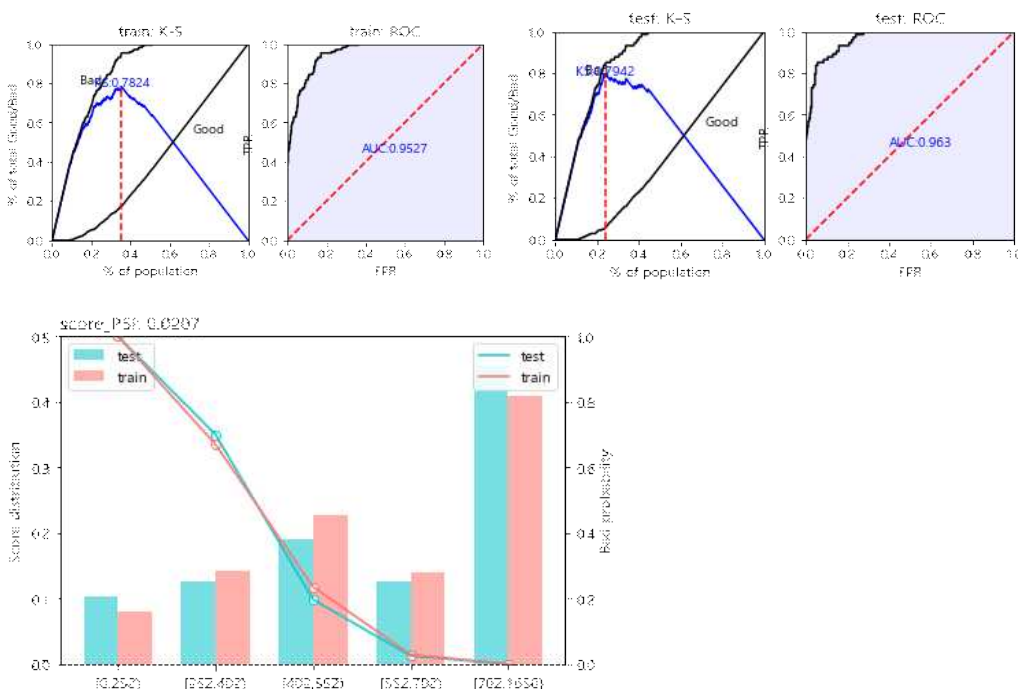
- 역전 현상 방지:

각 등급마다 신용 위험도가 순차적으로 증가하도록 설정하여 역전 현상을 방지하였습니다. 역전 현상이란 예상과 반대로 높은 점수 그룹의 불량률이 낮은 점수 그룹보다 높아지는 현상입니다.

따라서 본 연구에서 개발한 스코어 카드와 등급 체계를 통해 대구 광역시 읍면동에 거주하는 고객들의 신용 위험도를 정확하게 평가할 수 있게 되었습니다. 이러한 결과는 금융 기관들이 보다 효과적인 신용 평가 및 관리를 수행하는데 크게 도움을 줍니다.

III. 분석결과

○ 등급 체계 모형 성능 및 안정성 평가



- 본 연구에서 개발한 스코어 평점 모형은 학습 데이터의 KS 통계량이 0.7824, 검증 데이터의 KS 통계량이 0.7942로 우수한 성능을 보였습니다. 또한, 전체 모형에 대한 PSI(Population Stability Index)는 0.0207로 나타났습니다.
 - 우수한 구분력: KS 통계량은 예측 모형이 두 그룹을 얼마나 잘 구분하는지를 측정하는 지표입니다. 본 연구에서 개발된 모형의 KS 통계량이 높게 나타남으로써, 이 모형이 '소멸 위험'과 '비소멸 위험' 그룹을 잘 구분하고 있음을 확인할 수 있습니다.
 - 안정적인 성능: PSI는 학습 데이터와 신규 데이터 간의 점수 분포 차이를 측정하는 지표로, 값이 작을수록 좋습니다(일반적으로 PSI가 0.1 이하면 안정적으로 판단). 본 연구에서 개발된 모형의 PSI가 매우 낮게 나타남으로써, 이 모형이 시간 경과와 함께 변화하는 환경에도 안정적인 성능을 유지할 것임을 보여줍니다.
- 따라서 본 연구에서 개발한 스코어 평점화 모형은 우수한 예측 성능과 안정성을 동시에 가지고 있으며, 이는 대구 광역시 소멸 위험이라는 중요한 사회 문제를 해결하는데 크게 기여할 것입니다.

○ 대구 광역시만의 소멸 지역 위험 평점화 모형 도출

위험지역 평점화 모형 요약

No	항목명	구간	건수	구성비	비위험	위험	소멸위험률	woe	bin_iv	total_iv	breaks	is_special_value	points
1	MAN_SALE_AMT_20G	[-Inf, 4552]	206	90.0%	201	5	2.4%	-1.1705	0.7655	2.4827	4552	False	143
	MAN_SALE_AMT_20G	(4552, Inf]	23	10.0%	11	12	52.2%	2.6104	1.7072	2.4727	Inf	False	-318
⋮													
143	MAN_SALE_AMT_60GU	[-inf,200000.0]	91	0.085526316	30	61	0.67032967	1.900662091	0.39762428	0.890997889	1500000000	FALSE	52
	MAN_SALE_AMT_60GU	[200000.0,3200000.0]	54	0.05075188	26	28	0.518518519	1.265093581	0.102523792	0.890997889	2000000000	FALSE	0
	MAN_SALE_AMT_60GU	[3200000.0,inf]	490	0.460526316	368	122	0.248979592	0.086923715	0.003559969	0.890997889	6500000000	FALSE	-39

*adj_bin_values.csv, card_point.csv [신용평가모형 구축 폴더]

- 본 연구에서는 총 28개의 독립변수를 활용하여 대구 광역시만의 소멸 지역 위험 평점화 모형을 개발하였습니다. 이 모형은 지역 사회의 소멸 위험을 정량적으로 평가하고, 이를 관리하는 데에 중요한 도구로 작용합니다.

<장점 및 차별성>

- 지역 특성 반영: 본 모형은 대구 광역시의 특정한 사회경제적 특성과 문제를 반영하는 28개의 독립변수를 사용하였습니다. 따라서 이 모형은 대구 광역시에 맞춤형된, 고도로 적합한 소멸 위험 평가 도구입니다.
- 다양한 변수 활용: 다양한 분야에서 추출된 여러 변수들을 사용함으로써, 복잡하고 다양한 요인들이 지역 사회의 소멸 위험에 어떻게 영향을 미치는지 보다 정확하게 파악할 수 있습니다.
- 정책 수립 지원: 이 모형은 지역사회 개발 및 복지 정책 등을 수립하는 데 필요한

중요 정보를 제공합니다. 각 지역의 소멸 위험 점수는 해당 지역에 필요한 서비스와 자원 배분 등에 대해 중요한 참고 자료가 됩니다.

- 지속적인 업데이트 가능성: 본 모형은 새로운 데이터가 제공될 때마다 업데이트할 수 있으므로, 시간이 변함에 따른 사회경제적 변화를 반영하여 보다 실질적인 평가 결과를 제공할 수 있습니다.

따라서, 이 연구에서 개발된 대구 광역시 전용 소멸 위험 평점화 모형은 그 자체로 큰 가치를 가지고 있으며, 현재와 미래의 정책 결정 과정에서 유용하게 활용될 것입니다.

IV. 활용방안

○ Tableau 대시보드 구축 및 소멸 위험에 대한 정책 제언



https://public.tableau.com/app/profile/.47111185/viz/_16935562589520/2?publish=ye

도시 내의 미활용 공간을 활용하여 지역 경제에 활력을 불어넣는 새로운 방안을 제시하고자 합니다. 이 프로젝트는 도시의 짜투리 공간에 소규모 텃밭을 설치함으로써, 도심에서도 신선한 농수산물을 생산하는 것이 가능함을 보여주하고자 합니다.

이러한 접근법은 지역 사회에 다양한 이점을 제공합니다. 첫째, 이것은 신선하고 건강한 농수산물의 접근성을 향상시킵니다. 도시 주민들은 그들 자신의 커뮤니티에서 생산된 식품을 구입할 수 있게 되며, 이는 지역 내에서의 식료품 마일리지를 줄여 환경에 대한 부담을 감소시킵니다.

둘째, 소규모 텃밭은 도심 내에서 일자리를 창출하며, 이는 지역 경제를 활성화하는 데 기여합니다. 그 결과 현지 주민들에게는 취업 기회가 생기며, 전체적으로 동네가 번영하게 됩니다.

셋째, 도심 중심부에서의 장소 별 농장 운영은 커뮤니티 구성원들이 자신들의 음식 공급에 대해 직접적인 책임감과 연결감을 가질 수 있는 기회를 제공합니다.

마지막으로, 텃밭 설정은 그린 스페이스와 친환경적인 인구 지속 가능성 증진 사례로서 유용할 수 있습니다.

따라서 우리 프로젝트는 지속 가능한 도시 개발 모델 및 혁신적인 경제 활력 회복 전략으로서 중요합니다.

V. 기타

1. <https://skyeong.net/30>: 기계학습 기반의 신용평가 모형 개발과 신용 점수 계산
2. https://www.niceinfo.co.kr/creditrating/bi_score_4.nice :성능지표 산출원리
3. <https://certlys82.tistory.com/entry/부동산-유망지역-청년은-어디로-이동하는가-청년층-인구-이동-패턴> : 부동산 유망지역, 청년은 어디로 이동하는가? 청년층 인구 이동 패턴
4. <https://dacon.io/competitions/official/236088/codeshare/8305> : "신용평가모형"을 활용한 설명 가능한 성인병 보험료 등급 책정 아이디어!
5. https://choonghyunryu.github.io/posts/2021-02-07-optimal_binning/ : 스코어카드와 최적 비닝
6. <https://youtu.be/lKzyV9A2Oho> : [최강1교시] 수도권 쏠림과 지방위기 ep.01 | 도시계획 전문가
7. <https://youtu.be/DBGF50tw2DM> : 한국 도시들 이대로 괜찮은 걸까? 도시에 대한 궁금증들 총정리! 지방소멸, 한국 간판, 세종시, 서울로 7017
8. <http://www.pn.or.kr/news/articleView.html?idxno=13672>: "청년이 떠나는 도시"... 2030년 대구의 미래는?
9. <https://terms.naver.com/entry.naver?docId=5867835&cid=43667&categoryId=43667>: 시사상식사전-지방소멸위험지수
10. <https://velog.io/@macang15/공부로그-태블로-대시보드-따라하기-인구소멸-위험지도>: [공부로그] 태블로 대시보드 따라하기 - 인구소멸 위험지도
11. [소상공인 신용평가모형 구축에 관한 연구]
12. [빅데이터분석기법을이용한소상공인신용평가모형구축연구]