

지도 및 예측 모델을 활용한 전 국민 금융 스타일 시각화

[부동산 가격과 대출 트렌드 상관관계 조사]

방은지·이세현·전기영

1 Credit card data 분석

1. 인터랙티브 지도를 통해 알아본 전 국민 금융 스타일
2. 예측모델을 활용한 신용도에 중요한 요소 조사(전국)
3. 부동산 가격과 대출 트렌드 상관관계 조사

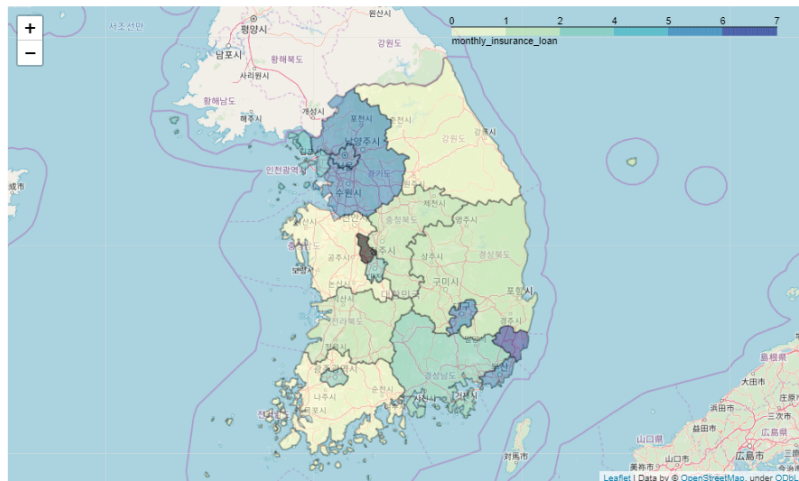
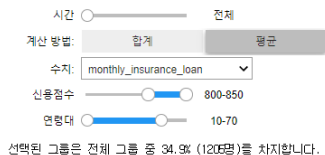
2 Jeju finance life data 분석

1. 인터랙티브 지도를 통해 알아본 제주 지역 금융 스타일
2. 예측모델을 활용한 신용도에 중요한 요소 조사(제주)

1. 인터랙티브 지도를 통해 알아본 전 국민 금융 스타일

1. Credit card data descriptions

- 시간(2016.01~2017.12), 계산 방법(합계, 평균), 수치(주어진 데이터), 신용점수(720~850), 연령대(10~90대)에 따른 인사이트를 한 눈에 알아볼 수 있도록 그래프 생성



인터랙티브 영상 주소: <https://youtu.be/qKWpepe4cpU>

2. 예측모델을 활용한 신용도에 중요한 요소 조사(전국)

1. Credit card data descriptions

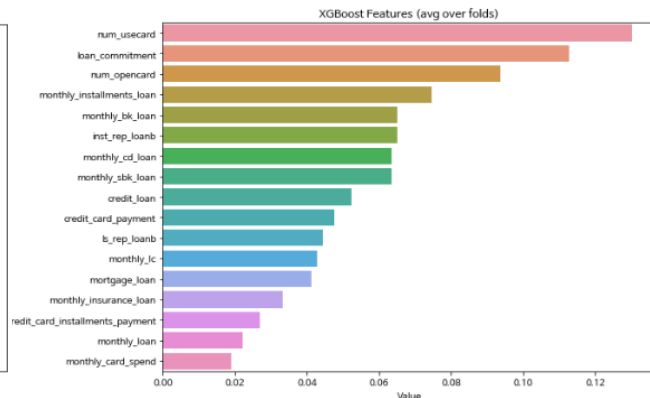
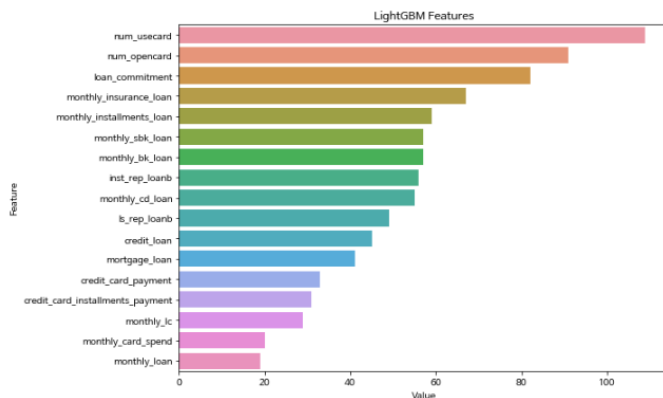
1. 신용점수 예측모델 학습 및 분석

신용점수를 예측하는 모델을 학습하여 신용도에 영향을 많이 주는 요소를 분석한다.

- 예측 목표: 평균 신용점수
- 학습에 사용되는 요소: 카드 사용량 관련 정보(카드 보유/사용 갯수, 카드 지출 등), 대출 관련 정보(총 대출 금액, 대출 잔액, 주택 담보 대출 금액 등)
- 학습에서 제외된 요소: 구분코드와 관련된 정보(성별, 도시), 연/월, 인구수(데이터 수가 적어 학습에 bias가 생기는 것을 우려), 평균 신용등급(예측 목표의 정보를 담고있음)

2. 변수 중요도 분석

- Feature Importance: 해당 요소가 데이터를 신용도 예측에 얼마나 큰 영향이 있는 지 분석

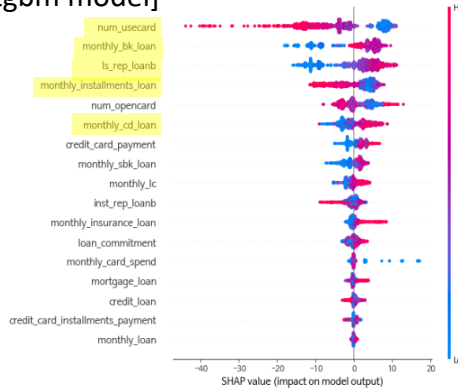


2. 예측모델을 활용한 신용도에 중요한 요소 조사(전국)

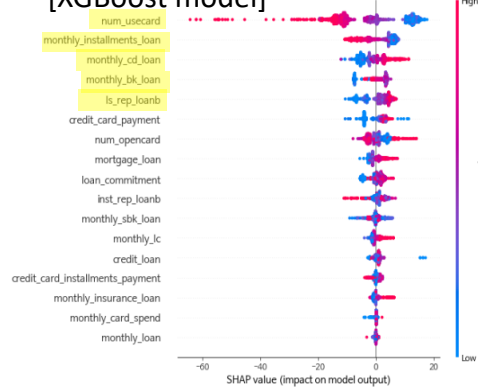
1. Credit card data descriptions

- 위의 중요도로는 변수가 얼마나 영향을 주지만 알 수 있을 뿐, 신용도를 높이거나 낮추는 데에는 어떤 영향이 있는지 알 수 없다.
- SHAP를 사용해 변수의 값이 높을/낮을 때 신용도가 올라가는지 내려가는지를 보았다.
- 공통된 top5 변수를 노란색으로 표시함

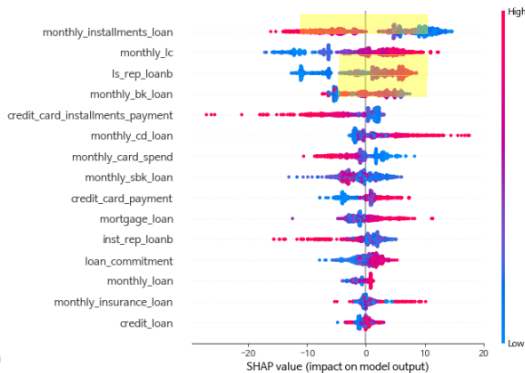
[Lightgbm model]



[XGBoost model]



- SHAP 분석에 따르면 사용 카드가 많을수록 신용도가 낮고 적을수록 높다고 예측한다.
- 하지만 사용카드의 수가 너무 큰 영향을 끼치는 것으로 보이기에 'num_usecard'와 'num_opencard'를 제외하고 재학습.



나머지 변수들을 토대로 신용도를 추정하자면 비례관계가 불확실한 (값의 높고 낮음이 혼재된) 변수를 제외하면

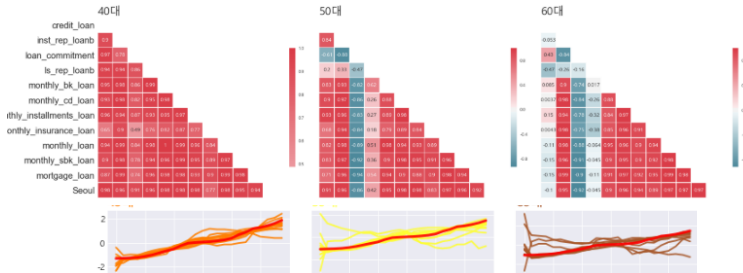
- 할부금융 대출이 적을수록
- 일시상환대출 잔여금액이 많을수록
- 월간 카드 사용량이 적을수록
- 월간 카드 대출량이 많을수록
- 신용카드 사용금액이 적을수록
- 대출금액 총합이 많을수록

신용도가 높게 추정된다.

3. 부동산 가격과 대출 트렌드 상관관계 조사

1. Credit card data descriptions

서울지역 연령대 별 대출트렌드 및 해당 지역 주택가격간의 상관 관계



1. 가정

- 한국인은 '내 집 마련'에 대한 수요가 강함.('내집을 꼭 마련해야 한다'는 의사 비중이 82.5% 달함)
- 축적된 자산이 없는 청년, 사회초년생, 신혼부부의 집값 기준 대출 부담이 훨씬 더 컸음.
- 따라서 우리는 대출 트렌드 (특히 30,40대의 대출금액)와 주택가격에 양의 상관관계가 있다고 가정함.

2. 과정

담보가치를 산정할 때, 금융감독원에선 한국감정원, KB국민은행을 포함한 4곳의 부동산 시세가 공신력 있다고 보며, 현재 규정에는 그 중에 하나로 하게 돼 있음. (따라서, 우리는 KB부동산(<https://hoy.kr/8IL2C>) 데이터를 주택가격의 기준으로 삼음.)

(1)Line plot

- 'x축: 시간(2016.1월~2017.12월), y축: 정규화된 11개의 컬럼 및 지역별 주택 가격 '
- 시간에 따라 데이터가 어떻게 변화하고, 지역별/연령별 대출 데이터와 주택가격의 트렌드를 보고자 함

(2)Heatmap

- 11개 컬럼 및 지역별 주택가격 데이터와의 상관관계
- ' 흰색: 기준점, 0 '
- ' 붉은색: 양의 상관관계, 0 ~ 1 ', 강한 양의 상관관계 0.5 이상
- ' 푸른색: 음의 상관관계, -1 ~ 0 ', 강한 음의 상관관계 -0.5 이하

3. 결과

대도시일수록, 30,40대 일수록 대출트렌드와 주택가격은 양의 상관관계를 보임.

-서울,인천,대전,광주,부산,울산,경기,강원,전북,전남,제주: 양의 상관관계

☞경제활동 유무도 대출심사 시 영향을 끼친다고 가정 할 때, 노년층(만 65세 이상 이므로, 60대는 제외함)의 경제활동 참여율은 부산, 울산을 포함한 6개의 도시가 타 도시에 비해 높을 수 있다.

☞ 또한, 대출 간의 상관관계가 높은 것으로 보아, 대출 실행자 한 명당 여러 대출상품을 소지할 수 도 있음이 보임

4. 따라서 해당 데이터들을 통해 **인덱스 형태의 서비스**를 제공함으로써 소비자 편의성이 대폭 올라갈 것으로 기대된다.

1. 인터랙티브 지도를 통해 알아본 제주 지역 금융 스타일

2. Jeju finance life data descriptions

- 수치에 있는 'job_' 항목 선택시 해당 직업을 가진 주민들의 거주 지역이 표시됩니다.
- 직업에 있는 항목을 선택하면 해당 직업을 가진 데이터를 선택합니다.

지도 배경: 기본 ☒ 록 ☐ 록록

계산 방법: 합계 ☐ 평균 ☒

수치:

성별: 전체 ☒ 남 ☐ 여

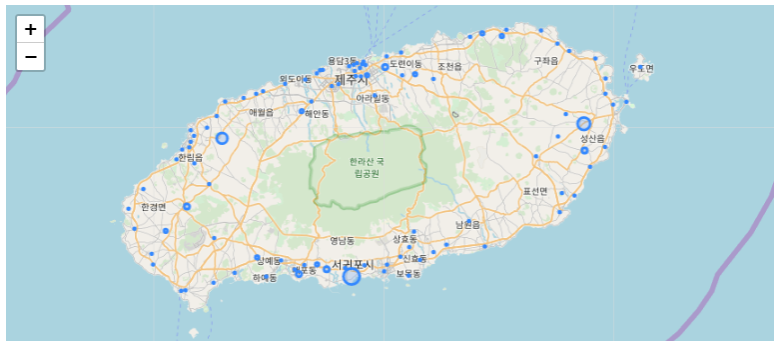
직업: 전체 ☐ 대기업 ☒ 중소기업 ☐ 공무원 ☐ 자영업

무직 ☐ 기타 ☐

신용등급: 4-6

연령대: 34-69

선택된 그룹은 전체 그룹 중 3.5% (366명)를 차지합니다.



활용 예시

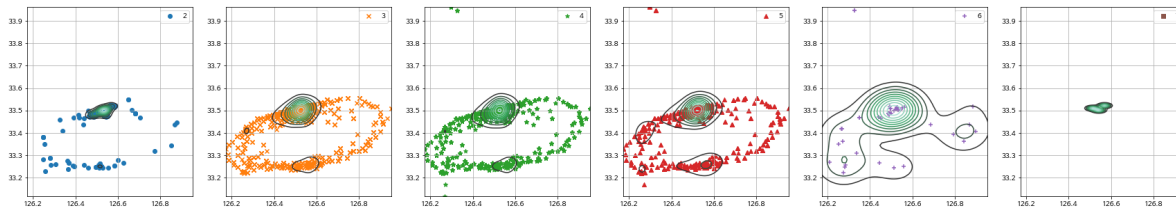
- 직업군에 따른 신용도 분포 확인
- 연령대별 평균 수입 확인
- 20대의 상위 신용등급의 평균 사용량 확인

인터랙티브 영상 주소: <https://youtu.be/Hw-ogRj8Hao>

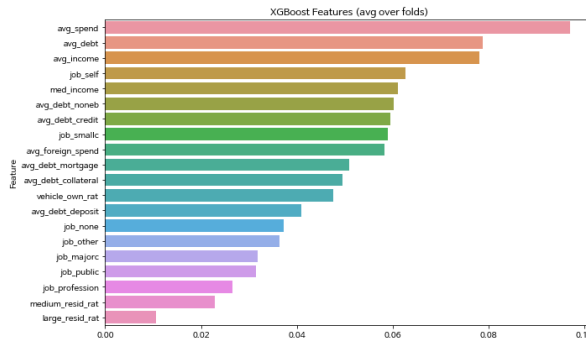
2. 예측모델을 활용한 신용도에 중요한 요소 조사(제주)

2. Jeju finance life data descriptions

등급별 분포



- 신용 등급별 분포를 보았을 때 구좌읍 근처에는 평균이 2등급인 지역이 없는 것으로 확인된다
- 데이터 중 가장 낮은 등급인 7등급은 제주시 주변의 도시에 거주한다.



학습/검증 데이터를 다양하게 분배하고 모델 파라미터를 여러가지로 조정해본 결과 공통적으로 중요하게 여겨지는 변수는

- 소비량(avg_spend)
- 소득 수준(med_income, avg_income)

중요치 않게 여겨지는 변수는

- 주택의 크기(large_resid_rat)
- 담보대출 중 일부(avg_debt_collateral, avg_debt_deposit)

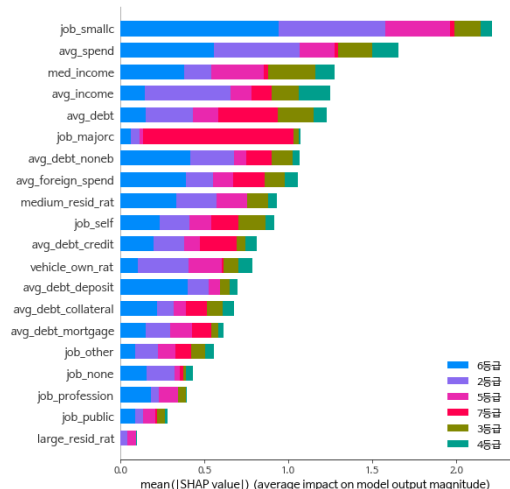
따라서 신용도에는 대출보다는 소비량과 소득의 영향이 크다고 할 수 있다.

2. 예측모델을 활용한 신용도에 중요한 요소 조사(제주)

2. Jeju finance life data descriptions

SHAP 중요도 분석

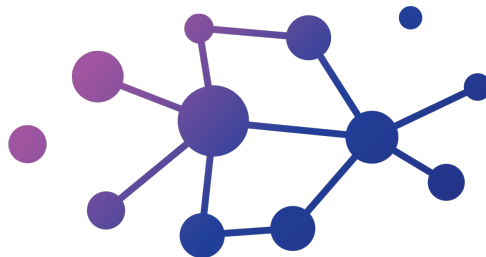
위의 중요도로는 변수가 얼마나 영향을 주는 지만 알 수 있을 뿐, 신용도를 높이거나 낮추는 데에는 어떤 영향이 있는지 알 수 없다. SHAP를 사용해 변수의 값이 높을/낮을 때 신용도가 올라가는지 내려가는지를 보인다.



각 등급에 가장 기여를 많이 하는 변수는

- 2등급: avg_income이지만 avg_spend, job_smallc와도 큰 차이가 없다.
- 3등급: med_income으로 역시 소득의 영향이 크다.
- 4등급: 영향이 미미해 구분이 어렵지만 avg_spend와 avg_income의 영향이 크다.
- 5등급: med_income으로 역시 소득의 영향이 크다.
- 6등급: job_smallc가 가장 영향이 크고, 그다음이 avg_spend이다.
- 7등급: job_majorc가 가장 영향이 큰 것이 놀랍다. 대기업에 종사하지만 7등급인 사람이 제주도에 많이 분포하는 것을 해석 가능하다.

THANK YOU



THANK YOU