#1 안녕하십니까 Untacted Banking service for Seniors, UBS의 발표를 맡은 백경동입니다. 혹시 발표 중 제 말이 너무 빠르거나 느리면 말씀해주시길 바랍니다. 저희는 시니어 타겟별 비대면 금융 서비스 교육 우선지역 파악을 위한 프로젝트를 진행하였는데요, 목차부터 보시죠.

#2 먼저 분석배경 및 목적, 활용한 데이터 및 타겟 정의 후 K-means와 Decision Tree를 활용한 분석과정 및 분석결과, 서비스 활용 방안, 기대효과 및 한계점 순서로 발표하겠습니다.

#3 분석 배경입니다. Untact란 contact와 un을 합성한 단어로, 접촉없이 서비스를 이용하는 경향을 의미합니다. 금융권에서 이러한 언택트 서비스의 개발 및 이용이 늘어나고 있고 코로나로 인해 더 빨라지고 있습니다. 이러한 경향에 따라, 여러가지 비용절감을 위해 은행은 점포를 줄이고 있고 대면 서비스를 주로 이용하는 노년층의 불만이 늘어나고 있습니다.

#4 반면, 한국은 인구의 고령화가 빠르게 진행되고 있습니다. 고령인구 비율이 일본을 넘어설 거라는 예측도 있습니다. 다음으로, 연령대별 모바일 뱅킹 이용률을 보면 60대 이상의 이용률이 크게 저조한 것을 알 수 있습니다. 이에 디지털 격차라는 신조어까지 나올 정도입니다. 또한 이 그래프에서 주목해야 할 점은 현재 50대인 베이비붐 세대의 데이터입니다. 60대로 접어들 이들의 경우, 기존의 고령층에 대한 인식과 다르게 어느정도 모바일뱅킹을 이용하며 비대면 서비스에 대해 거부감이 적습니다. 또한, 현재 모바일 금융 서비스를 주로 이용하는 젊은 층보다 소득이 높고 연금과 같은 고정적인 수입이 있기 때문에 잠재고객으로서 공략한다면 저금리 시대, 은행의 안정적인 장기고객으로 발전할 가능성이 클 것으로 보입니다.

#5 저희는 따라서 노년층을 디지털 격차를 해소하는 사회공헌적 목적을 가진 타겟A와 잠재고객 유치를 통해 은행의 이익을 창출하기 위한 타겟 B로 나누어 밀집 지역을 파악하고 은행분포와 비교하여 비대면 금융 서비스 교육 우선 지역에 대해 파악해보고자 합니다.

#6 활용한 데이터에 대한 정의입니다. 타겟별로 각각 3개의 데이터를 통해 밀집지역을 파악하였습니다. 사회취약계층은 기초생활수급자, 독거노인, 요양기관 정원 데이터를 이용하였습니다. 잠재고객층은 생활금융, 국민연금, 생활인구 데이터를 이용하였습니다. 저희가 행정동별 소득을 파악하기 위해서 서울시 빅데이터 캠퍼스에 방문하여 생활금융 데이터를 이용하였는데요, 보안상의 이유로 반출이 불가능하여 제출한 시연영상에서는 제외되었는데, 이미지를 캡쳐해 반출하여 발표 슬라이드에 사용하였음을 알려드립니다. 은행의 분포 파악에는 서울시 금융기관 데이터를 통해 다음의 시중은행들의 분포만을 이용하였습니다.

#7 그럼 사회취약계층의 요인별 시각화 자료를 먼저 보시겠습니다. 기초생활 수급자의 경우 등촌3동, 중계2,3동 등에 많았습니다. 독거노인의 경우 역시 강북지역과 천호2동, 등촌3동 등에 많이 분포하였습니다. 요양기관은 행정동이 아닌 구 데이터인데요, 도봉 노원 등 강북지역과 강서구, 은평구 등 강서지역에 요양기관이 많이 분포하였습니다.

#8 각 요인별 상위 10개지역입니다.

#9 다음으로 잠재고객층의 요인별 시각화 자료를 보시겠습니다. 평균 소득의 경우 역시 강남의

역삼1동, 신림동 등에서 높게 나타났습니다. 연금은 구 데이터인데, 역시 강남구와 서초, 송파구 등에서 높게 나타났습니다. 유동인구의 경우 여의동과 종로1,2,3,4동, 역삼동에서 높게 나타났습니다.

#10 각 요인별 상위 10개 지역입니다.

#11 다음으로 은행의 분포입니다. 앞서 분석 목표에서 말씀 드렸듯, 비대면 금융 서비스 교육이 필요한 지역은 은행의 분포가 적어 거동이 불편한 노인들이 대면 서비스를 위해 타 지역으로 이동해야만 하는 불편함이 있는 지역입니다. 따라서 상위 10개지역이 아닌 하위 지역을 파악해야 합니다. 분포 하위 지역 중 몇몇 지역은 데이터 상 은행 분포가 0인 지역도 있었습니다. (html 띄워 보여주며) 종로구 창신3동, 용산구 후암동 등 은행이 없는 지역은 금융 서비스를 대면으로 이용하기 위해서는 타 지역으로 이동해야 하는 불편함이 있었습니다.

#12 그렇다면 교육 우선 지역 파악을 위한 데이터 처리방안에 대해 말씀드리겠습니다. 먼저, 데이터들이 단위가 다 다르기 때문에 Min-Max Sacling을 통해 정규화하여 사용하였습니다. 앞서 보셨던 은행의 분포는 데이터가 적은 쪽이 1이 되도록 설정해야 하기 때문에 1에서 정규화한 값을 빼서 보정하여 사용하였습니다. 정규화한 값들을 각각 가중치를 곱하여 합산 후 각 타겟의 밀집지역 파악에 사용하였습니다. 가중치에 대한 설명은 다음 슬라이드에서 이어서 하겠습니다. 위 과정을 통해 K-means를 사용하여 군집화하였습니다. 군집화한 결과를 통해 우선 지역을 선정하고, 이를 타겟으로 의사결정나무를 활용하여 모델링하였습니다.

#13 가중치에 대한 설명입니다. 여러 논문과 연구 보고서 등을 참고하여 각 타겟별로 이용한 변수들 중 저소득층과 소득을 중요하게 보고 다른 변수들보다 높은 가중치를 주었고, 이러한 가중치를 줬을 때 k-means와 decision tree 모델이 가장 좋은 정확도를 보였습니다.

#14 각각의 가중치를 곱한 값을 이용하여 타겟별로 k-means를 이용하여 군집화하였습니다. K-means는 그룹간 거리와 비유사도 등의 비용함수를 최소화하는 방식으로 군집화가 이루어지는데, 이 때 클러스터의 갯수에 따라 성능이 크게 달라집니다. 저희는 Elbow Method를 이용하여 기울기가 완만해 지는 Elbow point를 적정한 K값으로 판단하였습니다. 타겟A와 B 모두 3~4개의 Cluster가 적정 갯수로 판단할 수 있습니다. 3개와 4개 중 더 적절한 갯수에 대해 다음 슬라이드에서 파악해보겠습니다.

#15 먼저 사회취약계층의 경우 실루엣 메소드를 이용하여 실루엣 점수 비교를 하였습니다. 비슷한 값을 가지지만 3개일 때가 4개일 때보다 좀 더 좋은 값을 가졌습니다. 3개 중 이 부분이 타겟이 밀집하여 있으면서 은행의 분포가 적은 교육 우선지역입니다. 이처럼 (1,1)에 가까울 수록 교육 우선순위가 높은 것을 보이기 위해 앞서 은행의 분포를 1- 정규화한 값을 빼서 보정하였습니다.

#16 잠재고객층의 경우 3개보다는 4개일 때 좀 더 좋은 실루엣 점수를 가졌습니다. 4개 중 이 부분이 잠재고객 타겟이 밀집하여 있으면서 은행의 분포가 적은 교육 우선지역입니다.

#17 아까 보여드렸던 그래프 속 Cluster의 중심점 좌표들입니다. 중심점이 (1,1)에 가장 가까운 cluster3와 cluster4가 각각 교육 우선지역이며 이 지역들은 타겟이 많고 은행이 적은 지역들입니다.

#18 군집화한 결과를 가지고 의사결정나무 모델링을 해보았습니다. 교육 우선지역인 cluster를 1로, 나머지 지역을 0으로 라벨링하여 의사결정나무의 target 변수로 사용하였습니다.

전체 데이터를 8:2로 분리한 후 train 데이터로 모델링, test 데이터로 모형의 정확도를 계산하였습니다. 모델링한 결과를 통해 서울과 비슷한 특성을 가진 대도시에서도 비대면 교육 우선 순위 지역을 파악할 때 수치적인 근거로 사용할 수 있을 것입니다.

#19 사회취약계층의 Decision Tree 결과입니다. 요양기관 정원이 0.621보다 작다의 False, 은행 합계가 0.653보다 크다의 False일 때 우선지역으로 분류되었습니다. 즉, 요양기관 정원이 0.621보다 크고 은행합계가 0.653보다 작은 지역이 교육 우선지역입니다. 모형의 정확도는 약 74.8%입니다.

#20 잠재고객층의 Decision Tree 결과입니다. 은행합계가 0.217보다 작다의 True, 연금평균이 0.825보다 크다의 True 때 우선지역으로 분류되었습니다. 즉 은행합계가 0.217보다 작고 연금평균이 0.825보다 큰 지역이 교육 우선지역입니다. 모형의 정확도는 약 68.4%입니다.

#21 Decision Tree의 분석 결과를 요약해보겠습니다. 요양기관 정원이 많고 은행분포가 적은지역, 은행분포가 적고 연금 수령액이 큰 지역이 우선 지역일 것으로 예측됩니다. 하지만 각각의 결과가 명확하게 분류되었다기보다 엔트로피가 꽤 큰 값을 가지고 정확도도 90%보다 낮습니다. 즉, 더 정확한 모형을 위해서는 추가적인 정보가 필요할 것입니다.

#22 위의 분석과정들을 통해 종합적인 비대면 금융 서비스 교육 우선 지역 선정 결과를 도출하고 시각화해보았습니다. 사회취약계층의 비대면 교육우선 지역은 진한 녹색의 cluster 3입니다. 취약계층 노인이 많고 은행의 분포가 적은 지역입니다.

#23 인구가 많은 노원구, 도봉구, 중랑구, 강북구 일대와 요양 기관이 많았던 은평구, 은행분포가 적었던 강동 일대와 동작구 등이 사회취약계층을 대상으로 한 비대면 금융 교육 우선지역입니다.

#24 잠재고객층의 분석 결과를 시각화해보았습니다. 진한 녹색의 cluster4가 교육 우선 지역입니다. 잠재고객층이 많고 은행의 분포가 적은 지역입니다.

#25 소득이 높은 강남3구, 인구 밀도가 높은 노원구, 은평구, 강서구 등이 잠재고객 유치를 위한 비대면 금융 교육 우선지역입니다. 노원구와 은평구는 사회취약층과 겹치는 것을 알 수 있습니다.

#26 서비스 활용방안입니다. 현재 노년층을 대상으로한 비대면교육이 없는 것은 아닙니다. 시중은행 및 지방 은행 등이 다양한 프로그램을 진행하고 있습니다. 하지만, 저희는 이러한 분석을 통해 타겟별로 더 필요한 교육을 더 효율적인 지역에서 진행하는데 사용하고자 합니다

.

#27 분석결과에서 보셨듯 취약계층과 잠재고객층의 우선 지역이 겹치는 곳도 있었지만 다른 곳도 있었습니다. 각각에 맞는 교육 프로그램을 설계하여 해당 우선 지역에서 진행한다면 더 효과적일 것입니다. 취약계층은 비대면 서비스 이용에 필요한 것들을 교육하는데, 잠재고객층은 편의를 제공하고 금융 상품 등 다양한 마케팅 전략과 함께 설계한다면 더 효과적일 것입니다.

또한, 실제 노년층을 대상으로 한 교육이 이루어지고 있기 때문에, 실제 교육 결과를 바탕으로 모형의 타겟 데이터를 보완한다면 더 좋은 성능의 모델을 구성할 수 있고, 서울 뿐만 아니라 다른 대도시에서도 적용 가능할 것입니다.

#28 기대효과입니다. 디지털 격차 해소의 사회공헌적 목적과 고객 유치의 목적을 달성할 수 있습니다. 대면 서비스의 고정비용을 줄이고 고객의 편의성과 만족도를 향상시킬 수 있습니다. 장기고객 전환 가능성이 높은 노년층 고객 유치에 효율성을 높일 수 있을 것입니다.

#29 한계점입니다. 군집분석을 통해 예측한 값을 타겟으로 모델링하였기 때문에 이중으로 예측을 한 것입니다. 모형에서 보여드렸던 정확도보다 실제로 정확도가 낮을 수 있습니다.

사회취약층의 경우 비대면에 필요한 pc나 스마트폰 보유 여부에 대해 알 수 없었습니다. 비대면 교육 서비스를 이용할 수 없는 사람에게 교육을 하는 것이 효율적이라고 볼 순 없겠죠? 취약계층의 경우 이 부분에 대한 보완이 필요합니다.

사용한 데이터 중 구 단위 데이터들은 해당 구의 행정동 갯수에 따라 평균을 내어 사용하였습니다. 이 때문에 약간의 약간의 오차가 발생할 수 있습니다.

#30 사용한 데이터들은 소득을 제외하고는 2018년 데이터로 통일하였습니다. 소득 데이터가 2019년인 이유는 생활금융 데이터가 2018년 9월~2019년 12월까지의 데이터였습니다. 2018년 4분기의 3개월만을 가지고 이용하기에는 어려움이 있어 2019년 데이터로 사용하였습니다.

#31 참고문헌들입니다.

#32 질문이 있으신 분들께 질문 받겠습니다. (질의응답 후) 감사합니다.