|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 2022년 NH투자증권 빅데이터 경진대회 (예선) | | | | | |
| **팀명** | Fine Bigdata Analysts (FBA) | | | | |
| **팀원** | **성명** | **생년월일** | **학교** | **학과** | **연락처** |
| 구본우 | 1999.04.12 | 한국과학기술원 | 산업 및 시스템공학과 | 010 9896 8258 |
| 최인수 | 1993.06.14. | 한국과학기술원 | 산업 및 시스템공학과 | 010 4021 7839 |
| 고우성 | 2000.07.30 | 연세대학교 | 경제학과 컴퓨터과학과 | 010 7304 3308 |
| **주제명** | Autoencoder 기반의 잠재 벡터를 바탕으로 한 고객 데이터 분류를 통한 페르소나 맞춤형 포트폴리오 최적화 서비스 | | | | |
|  | | | | | |
| **분석 및 개발 모형** | | | | | |
| * 기계학습 모델이 데이터 특징들의 관계를 정밀히 학습할 수 있게 체계적인 데이터가 공과 전처리를 실시하였으며 자세한 전처리 과정은 제출한 .ipynb 파일 내에 주석으로 보충 * 제공된 데이터 이외에도 월말 기준 고객 주식 잔고 정보 데이터를 바탕으로 고객 포트폴리오 위험 선호도를 측정하는 Value Weighted Volatility를 S&P Capital IQ 랑 investpy 패키지를 이용하여 추가 외부데이터로 사용 * 데이터들을 바로 군집화 하는 것이 아닌 잠재 변수 추출 및 차원 축소를 위해 Manifold Learning 의 Autoencoder와 비선형적 방법인 t-SNE를 사용 * Autoencoder는 Linear Autoencoder 와 Convolutional Autoencoder를 사용하였고 Autoencoder와 외부데이터 유무를 통해 총 6가지의 K-means clustering을 진행함.  1. 오토인코더와 외부데이터 둘 다 사용하지 않은 군집화 (Silhouette: 0.390) 2. 외부데이터를 포함하였지만 오토인코더를 거치지 않은 군집화 (Silhouette: 0.386) 3. 외부데이터 미포함 Linear Autoencoder 사용한 군집화 (Silhouette: 0.414) 4. 외부데이터 포함 Linear Autoencoder 사용한 군집화 (Silhouette: 0.415) 5. 외부데이터 미포함 Conv Autoencoder 사용한 군집화 (Silhouette: 0.583) 6. 외부데이터 포함 Conv Autoencoder 사용한 군집화 (Silhouette: 0.582)  * 위와 같은 6가지 군집 최적화를 실행 하였고 1번 실험의 실루엣 점수 0.390부터 6번 실험의 실루엣 점수 0.582까지 추가 외부데이터로 인한 성능 차이는 없었지만 오토인코더 유무의 따른 군집화 성능의 차이를 확인함. * Conv Autoencoder가 Linear Autoencoder보다 뛰어난 성능을 보였고 최적화 결과 데이터들은 Conv Autoencoder와 t-SNE를 이용해 2차원으로 축소하고 4개의 최적의 군집 수로 군집화를 실행 하였을 때 아래와 같은 시각화 결과가 나타남 * 분류된 각 고객 군집에 대한 통계량 분석에 따르면 주로 연령대와 수입으로 일차적으로 분류되어 향후 데이터가 추가될 경우 소비 등을 통해서 더욱 세분화된 분류가 가능할 것으로 판단됨 * 또한 추가 발전 가능성으로 **tot\_aet\_tld\_rnd\_202201** 와 같은 총자산 시계열 데이터의 변화를 분석 할 수 있는 오픈뱅킹 또는 계좌이체 데이터가 주어진다면 고객 총자산의 변화가 수익률에 따른 자산 변화인지 계좌이체에 따른 자산 변화인지 추가적으로 분석 가능할 것으로 판단됨 | | | | | |
| **(선택) 사용 외부데이터** | | | | | |
| 1. S&P Capital IQ 랑 investpy 패키지를 활용하여 Standard Deviation of Daily Log Normal Price Returns for Securities with Historic 3 Month Rolling Window Annualized by a Factor of 250. Data Sourced Based on Date: 2020-12-31 to Avoid Look-ahead Bias 방법을 활용 2. 포트폴리오 내 해당 증권의 비중을 계산해서 Value Weighted Volatility, 즉, Hypothetical 포트폴리오의 3달 과거 변동성을 추측함  * Value Weighted Volatility는 고객의 Risk-profile 또는 Risk-preference를 주어진 Feature보다 더 정확히 측정할 수 있다. 이미 존재하는 Feature는 증권 Type을 알려주지만 그 Type 내에서 얼마나 risk-averse 또는 risk-tolerant 한 지 시그널링 해주지 않는다. 우리는 volatility를 risk proxy로 설정해서 더 세밀한 투자 성향 인사이트를 뽑아낼 수 있을 것으로 보임 | | | | | |
| **서비스 기획 아이디어** | | | | | |
| * 본 제안서에서는 고객의 투자거래 성향을 분류하고 가계 금융 서비스의 특징과 연동한 포트폴리오 기반의 금융상품 추천 시스템을 기획하고자 함 * 특히 연령대 및 수입별로 달라지는 Target Date Fund 등과 같은 평생 자산 포트폴리오 관리(lifetime portfolio management) 기법을 고객의 투자 거래 성향과 결합하여 포트폴리오 운용을 위한 금융 상품 추천 시스템을 기획하고자 함 * 기본적으로 포트폴리오 서비스에 대해 평생 자산관리 최적화 관점에서 가계복지조사 등의 데이터를 참조하여 금융 포트폴리오 내에서 소비자의 이질성을 상세히 파악하고자 함 * 이러한 방식을 통해 기본적인 목표 투자 수익률 지표를 선정한 뒤 고객의 거래 성향 등을 파악하여 고객의 성향과 평생 자산 관리의 관점에서 고객의 목표 달성 가능성을 높임으로써 실패 가능성이 최소화되도록 투자 종목을 추천 * 이러한 방식이 채택될 때 고객의 데이터를 누적시키고 이외에 고객의 서비스에 대한 평가 등을 통하여 온라인 방식으로 고객의 투자 거래 성향을 기반으로 한 고객 네트워크 및 유사성 분석을 통하여 실제 수익률 등을 기반으로 더 신속하게 투자 포트폴리오를 제공받을 수 있는 종목 추천 서비스를 기획하고자 함 * 장기적인 관점에서 평생 자산 관리 서비스를 통해 고객의 충성심(loyalty)을 높일 수 있으며, 고객 이탈률에 대한 개선 가능성이 있으며 또한 서비스 형태 점점 개선될 수 있는 여지가 있어 시간에 비례하여 고품질의 서비스를 제공할 수 있다는 장점이 있음 * 또한 이론적 배경을 바탕으로 하는 것뿐만 아니라 군집화 및 고객 네트워크 등 시각화가 쉬운 방법론들을 활용하는 방식을 채택함으로써 설명 가능한(explainable) 서비스라는 점에서 금융 분야에서 최근 화두로 떠오르고 있는 설명 가능한 AI 금융 서비스라는 관점에서도 적합한 서비스로의 발전 가능성이 있음 | | | | | |