**프로젝트 기획안**

**2023년 9월 14일**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 프로젝트 조 | AI/BD트랙 2조 | | | |
| 프로젝트 팀원 | 팀명: 세렝게티  팀장: 원동환  팀원: 장성호, 최은정, 황지혜 | | | |
| 프로젝트 주제 | 강소기업 예측을 통한 우리은행 잠재고객 발굴 | | | |
| 프로젝트 주제 선정 배경 | 우리은행이 ‘기업금융 명가 재건을 위한 전략 발표회’를 열고, 2025년까지 기업대출 점유율 2위를 탈환하고 2027년까지 1위 자리에 오르겠다는 목표를 제시했다. 현재 우리은행은 4대 은행(KB국민·신한·우리·하나) 중 기업 대출 점유율이 가장 낮다. ‘기업금융 명가’라는 옛 명성이 무색하게 최근 기업대출 부문에서 맥을 못 추는 상황을 반전시키겠다는 포부로 전망된다.  또한 목표 달성을 위해 우리은행은 2027년까지 기업대출 규모를 237조9000억원으로 확대할 계획이라고 한다. 지난달 말 기준 우리은행 대출 자산 중 기업대출 잔액은 약 135조7000억원, 가계대출은 132조7000억원으로, 자산 비중으로 따졌을 땐 비율이 각각 50.5%와 49.5%로 엇비슷하다. 현재 5대5인 기업대출과 가계대출 비율을 기업대출을 더 확대하여 2027년까지 6대4로 조정하겠다는 것이다.  위와 같이 우리은행의 지속적인 기업대출 규모 증가 계획에 맞춰 본 프로젝트를 통해 중소기업 중 향후 발전 전망이 유망한 강소기업 예측을 통해 은행 기업고객 마케팅에 도움이 될 수 있을것으로 생각하였다.  강소기업이 가지고 있는 재무제표 및 평판 데이터들을 분석하여 각 요소들의 특징과 상관관계를 파악하여 예측 모델을 만들고 향후 우리은행의 기업고객 마케팅에 도움이 되도록 액션 플랜을 제안하고자 한다. | | | |
| 프로젝트 수행 방향  (주요 기능 설명) | **빅데이터** | | | |
| * 강소기업에 대한 정의 및 목표변수 수립 * 주제와 관련된 데이터 수집 (매출액, 평균연봉, 기업평점 등) * 재무제표와 기업평점 관련 데이터 추출하기 * 강소기업에 영향을 미치는 변수 추출하기 * 추출한 데이터로 가설 검증 및 상관관계 분석 및 시각화 * 주제 관련 머신러닝 모델 선택 * 머신러닝 학습 및 평가 진행 * 수집된 데이터에서 추출된 데이터 훈련 및 모델 구축 * 데이터 : DART, 공공데이터 포털, 잡플래닛, 원티드인사이트 등 | | | |
| 프로젝트 R&R  (조원별 R&R)  *\*추후 개인별  기여도 결과  평가 시 활용* | **원동환** | **장성호** | **최은정** | **황지혜** |
| 팀장  데이터수집  개발 | 데이터수집  개발 | 부팀장  데이터수집  개발 | 서기  데이터수집  개발 |
| 프로젝트 수행 도구 | **빅데이터** | **AI** |  |  |
| * Python * GitHub | * Colab * Python * Google Drive * jupyter notebook * GitHub |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 프로젝트 목표 | 빅데이터 분석을 통해서 데이터 전처리하여 강소기업의 패턴 및 이상치를 탐지하여 향후 유망한 강소기업을 예측한다.  재무제표 수집을 통한 기업의 건정성을 파악하여 강소기업을 발굴한다. | | | |
| 필수 기능 | **빅데이터** | **AI** |  |  |
| * 프로젝트 주제 관련 데이터 수집후 분석 결과 시각화 * 강소기업패턴분석 * 강소기업에 영향을 미치는 데이터 분석 | * 강소기업 데이터셋으로 훈련시킨 모델 개발 |  |  |
| 포함 기술 | **빅데이터** | **AI** |  |  |
| * Matplotlib, Seaborn 활용한 분석 결과 시각화 * 크롤링과 API를 활용한 데이터 수집 * 데이터 활용, 정형 및 비정형 데이터 처리 | * 강소기업 데이터 수집 및 정제 * Decision Tree알고리즘을 기반으로 한 분류 모델 구축 |  |  |

|  |  |  |
| --- | --- | --- |
| **주차별 수행계획** | **1주차** | **주제선정** |
| **2주차** | **주제 구체화** |
| **3주차** | **데이터 수집, 데이터 전처리** |
| **4주차** | **중간정검 , 데이터 전처리** |
| **5주차** | **데이터 수집 및 전처리** |
| **6주차** | **데이터 시각화** |
| **7주차** | **패턴 분석** |
| **8주차** | **지역별 비교 및 예측 모델 구축** |
| **9주차** | **예측모델구축** |
| **10주차** | **예측모델구축** |
| **11주차** | **분석결과예측** |
| **12주차** | **최종 산출물 발표 및 피드백 (수료식 전 점검)** |

**□ 기대효과**

현재 산재된 여러 중소기업의 요소들 (매출, 평판, 직원수 등) 을 분석하여 강소기업에 영향을 미치는 패턴을 분석하고

강소기업 예측모델을 제시하여 우리은행의 기업금융 규모 확대시 활용할 수 있는 유용한 자료를 제공할 수 있다.

위와 같은 인공지능 예측 모델을 통해 마케팅을 위한 인적 비용과 시간을 감소시킬 수 있고 타행보다 한 발 빠른 타겟팅 및

유망 강소기업 마케팅을 통해 우리은행 기업고객 확대에 기여할 수 있다.

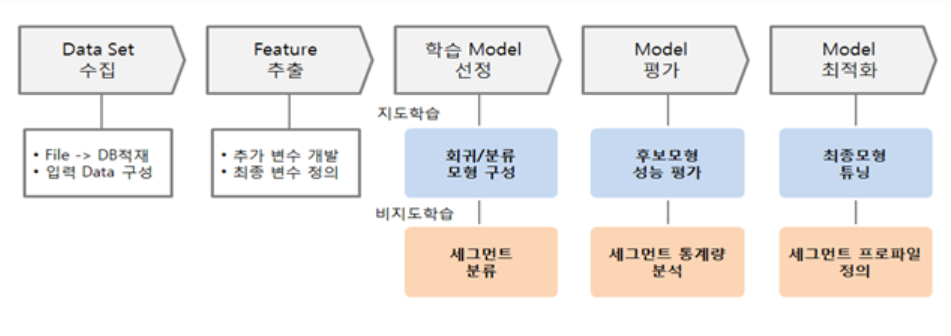
또한 지속적으로 강소기업 예측 서비스를 통해 중소기업시장을 모니터링하고 잠재고객을 꾸준히 탐색할 수 있다.

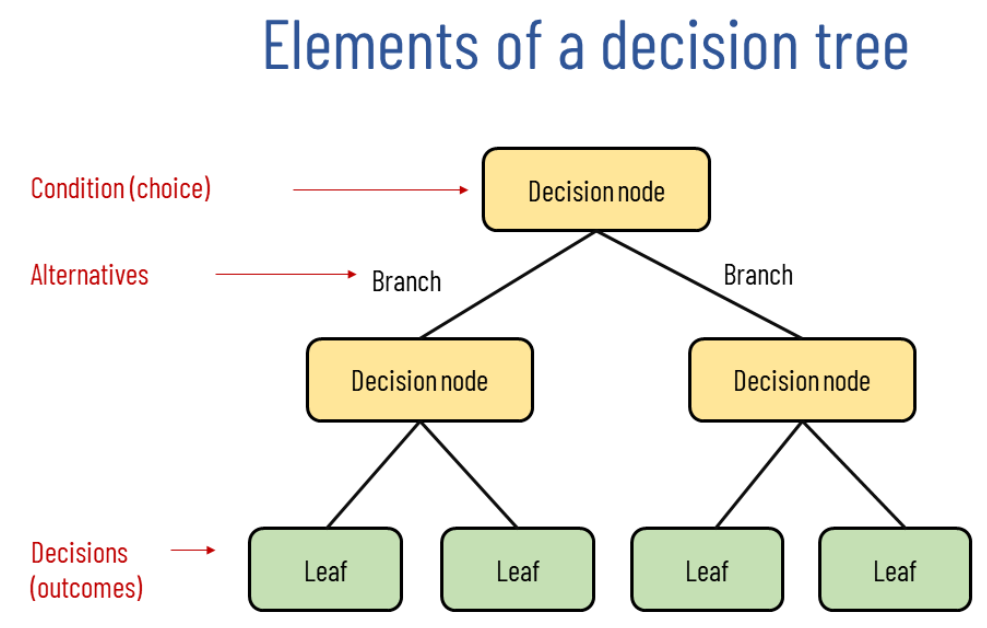
이를 통해 ‘우리은행의 2027년 까지 기업금융 명가 재건을 통한 1위 달성 목표’ 에 큰 보탬이 될 수 있을 것이다.

**□ WBS 첨부**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **업무** | **시작일** | **종료일** | **담당자** | **완료율** |  |  |  |  |  |  |  |  |  |  |  | ㈜ |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **1.프로젝트 개요** | 230825 | 230901 | 전체 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 프로젝트 목적 동기** | 230825 | 230901 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 프로젝트 범위** | 230825 | 230901 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **2. 데이터 수집 및 전처리** | 230901 | 230915 | 전체 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 강소기업 데이터수집** | 230901 | 230915 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 데이터 처리** | 230901 | 230915 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **3. 데이터 시각화** | 230916 | 231006 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 재무재표 내용** | 230916 | 231006 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 잡플래닛 평판 정보** | 230916 | 231006 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 상관관계** | 230916 | 231006 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **4. 패턴 분석** | 231007 | 231020 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 이상치 탐지를 통한 파악** | 231007 | 231020 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 강소기업에 영향을 미치는 요소 분석** | 231007 | 231020 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 다양한 변수를 고려한 강소기업 패턴 도출** | 231007 | 231020 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **5. 예측 모델 구축** | 231021 | 231110 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 강소기업비교분석** | 231021 | 231110 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 예측모델구축** | 231021 | 231110 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 모델 성능 평가 및 개선** | 231021 | 231110 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **6. 결과 해석** | 231117 | 231117 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 분석결과해석** | 231117 | 231117 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **- 강소기업 예측을 위한 제안** | 231117 | 231117 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**□ 분석 프로세스**





평가모델 : Decision Tree

**□ 예상 결과물 (분석 결과를 서비스로 개발 시 제안하고 싶은 프로토 타입)**

