

# 머신러닝 프로젝트 기획안

팀명	마당을 나온 수탉(Stock)
주제명	악성재고의 특징 분석 및 판매여부 예측

## 1. 역할분담

이름	역할
구경서	EDA, 데이터 시각화
김명진	분석 관련 자료 확보, 데이터 전처리
신민경	데이터 시각화, 발표 자료 제작
여혁수	모델링 알고리즘 제작, 발표
최은서	데이터 전처리, 알고리즘 제작

## 2. 주제 선정 배경

### 1. 산업적 배경

- 악성재고(Dead Stock)는 기업의 비용 부담과 수익성 저하를 유발하며, 특히 전자제품이나 패션 산업처럼 제품 수명 주기가 짧은 분야에서 수요 변화로 빠르게 진부화됨.
- 효율적인 재고 관리는 기업의 현금 흐름과 운영 효율성을 개선하는 중요한 요소이며, *Harvard Business Review*에 따르면 재고 회전을 개선은 기업 경쟁력을 높이는 핵심 전략임.

## 2. 데이터 활용의 중요성

- **데이터 기반 접근의 필요성**
  - 전통적으로는 경험이나 직관에 의존한 재고 관리가 이루어졌으나, 데이터 분석과 머신러닝 기술을 통해 보다 정밀한 예측이 가능해짐.
- **데이터셋의 특성 활용**
  - 판매량, 재고량, 할인율 등의 데이터를 분석함으로써 악성재고의 주요 특징을 파악할 수 있음.

## 3. 학문적/프로젝트적 배경

- **유의미한 인사이트 도출**
  - 악성재고를 정확히 정의하고, 이를 분석하여 도출된 특징은 기업의 의사결정에 큰 도움을 줄 수 있음.
- **예측 모델의 적용 가능성**
  - 머신러닝 알고리즘을 활용하여 악성재고의 판매 여부를 예측함으로써 실질적인 비즈니스 문제 해결에 기여할 수 있음.

## 4. 사회적/환경적 배경 (선택사항)

- **환경적 영향**
  - 악성재고의 처리는 폐기물 증가와 같은 환경문제를 초래할 수 있음.
  - 효율적인 관리로 불필요한 자원 낭비를 줄이는 데 기여할 수 있음.

## 3. 프로젝트 목표

### 1. 악성재고의 주요 특징 분석

- a. 판매량, 재고량, 정가, 할인율 등 다양한 변수와 악성재고 간의 상관관계를 분석하여 악성재고의 공통된 특징과 패턴을 도출.
- b. 기업이 악성재고를 사전에 식별하고 관리하는 데 도움을 줄 수 있는 실질적인 인사이트를 제공.

### 2. 판매 여부 예측 모델 개발

- a. 머신러닝 기반 예측 모델을 활용하여 악성재고의 최근 6개월간 판매 가능 여부를 예측.
- b. 다양한 모델의 성능을 평가하고, 예측 결과를 바탕으로 개선된 재고 관리 전략을 제안.

### 3. 비즈니스 의사결정 지원

악성재고를 줄이기 위해 프로모션, 재입고 조정 등 실질적인 대응 방안을 제안.

### 4. 활용 데이터

Historical Sales and Active Inventory

(<https://www.kaggle.com/datasets/flenderson/sales-analysis>)

## 5. 분석 방안

### 1. 데이터 탐색 및 전처리

#### 1. 결측치 처리

- NaN 값이나 잘못된 데이터를 확인하고 적절히 대체하거나 제거.
- 예: 평균값 대체, 0 으로 대체, 삭제 등.

#### 2. 데이터 시각화

- 변수 간 상관관계 분석 및 분포 확인.
- 주요 변수(재고량, 할인율, 판매량 등)와 악성재고 여부의 관계를 파악.
- Seaborn 이나 Matplotlib 를 사용하여 히트맵, 상자그림(Boxplot), 산점도 등을 생성.

#### 3. 이상치 탐지

극단적인 값(예: 너무 높은 재고량, 비정상적으로 낮은 실구매가 등)을 확인하고 처리.

#### 4. 변수 변환 및 파생 변수 생성

- 기존 변수에서 추가적인 파생 변수 생성:
  - 예: 재고 회전율 = 판매량 / 재고량, 할인율 = (정가 - 실구매가) / 정가.
- 이진 변수를 인코딩하여 활용.

### 2. 탐색적 데이터 분석 (EDA)

#### 악성재고 특징 도출

- 악성재고 여부와 변수들(재고량, 판매량, 정가, 할인율 등) 간의 관계 분석.

- b. 그룹별 평균, 중간값 등 기술 통계를 활용.
- c. 예: 악성재고의 평균 재고량과 일반 재고의 재고량 비교.

### 3. 판매 여부 예측 모델링

#### 1. 목표 정의

- a. 예측 목표: 악성재고의 최근 6개월간 판매 여부(1: 판매됨, 0: 판매 안 됨).

#### 2. 모델 선택 및 구축

- a. 주요 모델:
  - i. **로지스틱 회귀**: 간단하고 해석 가능한 이진 분류 모델.
  - ii. **랜덤 포레스트**: 비선형 관계를 잘 다루며 변수 중요도 제공.
  - iii. **XGBoost/LightGBM**: 복잡한 데이터에서도 높은 성능 제공.
  - iv. **KNN**: 가까운 데이터 간 유사성 기반 예측.
- b. 성능 비교를 위해 여러 모델을 사용하고 최적 모델 선정.

#### 3. 데이터 분할

- a. 데이터를 학습용/검증용/테스트용으로 분할.
- b. 예: 64% 학습, 16% 검증, 20% 테스트.

#### 4. 하이퍼파라미터 튜닝

GridSearchCV 사용해 모델 최적화.

#### 5. 평가 지표 설정

정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score 등을 활용.

### 4. 결과 분석 및 해석

#### 1. 변수 중요도 파악

- a. 모델에서 중요한 변수(feature importance)를 시각화하고 해석.
- b. 예: 할인율, 재고량, 고객 실구매가가 예측에 미치는 영향.

#### 2. 판매 여부와 주요 변수 간 상관관계 분석

- a. 판매 가능성에 영향을 미치는 변수의 패턴 도출.

#### 3. 의사결정 방안 제안

- a. 재고 관리 및 할인 정책 최적화를 위한 인사이트 제공.
- b. 예: 특정 할인율 이상에서는 판매 가능성이 높아지는 패턴 제안.

### 6. 결 론

## 1. 주요 결과 요약 (가정)

- 분석 결과
  - 악성재고의 특징: (예시) 높은 재고량, 낮은 할인율, 특정 마케팅 유형(예: D)에서 악성재고 발생 비율이 높음.
  - 예측 모델 성능: (예시) XGBoost 모델이 가장 높은 점수(예: 0.85)를 기록하며 판매 여부 예측에 적합한 성능을 보임.
- 핵심 인사이트
  - 할인율이 30% 이상일 때 판매 확률이 유의미하게 증가.
  - 특정 마케팅 유형(D)에서는 악성재고 발생률이 20% 이상 높아, 대체 전략 필요.

## 2. 비즈니스 적용 방안

- 재고 관리 전략
  - 악성재고 가능성이 높은 품목에 대해 사전 조치를 시행(할인율 조정, 재입고 중단 등).
  - 특정 변수(재고량, 할인율 등)를 고려한 프로모션 캠페인 설계.

## 3. 프로젝트의 한계

- 데이터 한계
  - 데이터가 특정 기간 또는 특정 제품군 기준으로 분류되어 있지 않고 과도하게 일반화되어있음.
  - 일부 변수(마케팅 유형, 실구매가 등)의 정의가 명확하지 않아 해석에 어려움이 있었음.
- 모델 한계
  - 판매 여부 예측의 정확도는 높지만, 실제 비즈니스 환경에서 적용할 때 고객 행동의 비예측 가능성 등 추가 요소 고려 필요.

## 4. 향후 과제

- 데이터 보강
  - 고객 프로필, 계절성 데이터 등 추가적인 변수를 수집하여 예측 모델 성능 개선.
  - 악성재고 발생 원인 분석을 위한 외부 데이터(시장 동향, 경쟁사 데이터 등) 통합.

- **모델 개선**

- 시계열 모델을 적용하여 재고 및 판매량의 변화 추세를 반영한 예측 가능성 탐색.
- 신경망 기반의 비선형 모델 활용으로 성능 향상 시도.

- **비즈니스 검증**

- 실제 기업 환경에서 모델 적용 후 성과를 측정하고 피드백을 반영한 개선 과정 필요.