

Bank_Marketing_Dataset_Info

데이터셋 특성

| 특성 | 설명 |
|--------|---------------------|
| 다변량 | 예 |
| 주제 분야 | 비즈니스 |
| 연관된 작업 | 분류 (Classification) |
| 특성 유형 | 범주형, 정수형 |
| 인스턴스 수 | 45,211개 |
| 특성 수 | 16개 |

데이터셋 정보

추가 정보

이 데이터는 포르투갈의 한 은행에서 진행한 직접 마케팅 캠페인과 관련이 있습니다. 마케팅 캠페인은 전화 통화를 통해 이루어졌습니다. 종종 같은 고객에게 여러 번 연락을 해야 했으며, 그 목표는 고객이 은행의 정기 예금 상품을 가입할지 ('yes') 아니면 가입하지 않을지 ('no') 예측하는 것이었습니다.

데이터셋은 총 네 가지로 구성되어 있습니다:

1. **bank-additional-full.csv**: 전체 예제(41,188개)와 20개의 입력 변수로 구성되어 있으며, 날짜별로 정렬되어 있습니다(2008년 5월에서 2010년 11월까지). [Moro et al., 2014]에서 분석한 데이터와 매우 유사합니다.
2. **bank-additional.csv**: 1)에서 무작위로 선택된 10%의 예제(4,119개)와 20개의 입력 변수로 구성됩니다.
3. **bank-full.csv**: 전체 예제와 17개의 입력 변수로 구성되어 있으며, 날짜별로 정렬된 더 오래된 버전입니다.
4. **bank.csv**: 3)에서 무작위로 선택된 10%의 예제와 17개의 입력 변수로 구성됩니다.

작은 데이터셋은 SVM과 같은 계산적으로 더 많은 리소스를 요구하는 머신러닝 알고리즘을 테스트하는 데 사용됩니다.

분류 목표는 고객이 은행의 정기 예금 상품을 가입할지 ('yes') 아니면 가입하지 않을지 ('no') 예측하는 것입니다.

변수 설명

| 변수명 | 역할 | 유형 | 인구통계학적 정보 | 설명 | 단위 | 결측치 |
|-----------|----|-----|-----------|---|----|-----|
| age | 특성 | 정수형 | 나이 | | | 없음 |
| job | 특성 | 범주형 | 직업 | 직업 유형 (범주형: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown') | | 없음 |
| marital | 특성 | 범주형 | 결혼 상태 | 결혼 상태 (범주형: 'divorced', 'married', 'single', 'unknown'; 'divorced'는 이혼 또는 과부를 의미함) | | 없음 |
| education | 특성 | 범주형 | 교육 수준 | 교육 수준 (범주형: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', | | 없음 |

| 변수명 | 역할 | 유형 | 인구통계학적 정보 | 설명 | 단위 | 결측치 |
|-------------|----|-----|-----------|---|----|-----|
| | | | | 'illiterate', 'professional.course', 'university.degree', 'unknown') | | |
| default | 특성 | 이진형 | | 신용 불량 여부 | | 없음 |
| balance | 특성 | 정수형 | | 연간 평균 잔고 | 유로 | 없음 |
| housing | 특성 | 이진형 | | 주택 대출 여부 | | 없음 |
| loan | 특성 | 이진형 | | 개인 대출 여부 | | 없음 |
| contact | 특성 | 범주형 | | 연락 방법 (범주형: 'cellular', 'telephone') | | 있음 |
| day_of_week | 특성 | 날짜형 | | 마지막 연락 요일 | | 없음 |
| month | 특성 | 날짜형 | | 마지막 연락 월 (범주형: 'jan', 'feb', 'mar', ..., 'nov', 'dec') | | 없음 |
| duration | 특성 | 정수형 | | 마지막 연락 지속 시간 (초 단위) | | 없음 |
| campaign | 특성 | 정수형 | | 이번 캠페인에서 고객에게 이루어진 연락 횟수 (마지막 연락 포함) | | 없음 |
| pdays | 특성 | 정수형 | | 이전 캠페인에서 마지막 연락 이후 지난 일수 (-1은 이전에 연락을 받은 적 없음) | | 있음 |
| previous | 특성 | 정수형 | | 이전 캠페인에서 고객에게 이루어진 연락 횟수 | | 없음 |
| poutcome | 특성 | 범주형 | | 이전 마케팅 캠페인의 결과 (범주형: 'failure', 'nonexistent', 'success') | | 있음 |
| y | 목표 | 이진형 | | 고객이 정기 예금을 가입했는지 여부 | | |

추가 변수 정보

입력 변수:

1. **age:** 나이 (숫자형)
2. **job:** 직업 (범주형: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
3. **marital:** 결혼 상태 (범주형: "married", "divorced", "single", "divorced"는 이혼 또는 과부를 의미)
4. **education:** 교육 수준 (범주형: "unknown", "secondary", "primary", "tertiary")
5. **default:** 신용 불량 여부 (이진형: "yes", "no")
6. **balance:** 연간 평균 잔고 (숫자형)
7. **housing:** 주택 대출 여부 (이진형: "yes", "no")
8. **loan:** 개인 대출 여부 (이진형: "yes", "no")
9. **contact:** 연락 방법 (범주형: "unknown", "telephone", "cellular")
10. **day:** 마지막 연락일 (숫자형)
11. **month:** 마지막 연락 월 (범주형: "jan", "feb", "mar", ..., "nov", "dec")
12. **duration:** 마지막 연락 지속 시간 (초 단위, 숫자형)

- 13. **campaign:** 이번 캠페인에서 이루어진 연락 횟수 (숫자형)
 - 14. **pdays:** 이전 캠페인에서 마지막 연락 이후 지난 일수 (숫자형, -1은 이전에 연락을 받은 적 없음)
 - 15. **previous:** 이전 캠페인에서 이루어진 연락 횟수 (숫자형)
 - 16. **poutcome:** 이전 마케팅 캠페인의 결과 (범주형: "unknown", "other", "failure", "success")
- 목표 변수 (예측 대상):
- 17.
 - y:** 고객이 정기 예금을 가입했는지 여부 (이진형: "yes", "no")

출처: UCI Bank Marketing Dataset