**빅데이터 이해 3주차 과제**

학번: 20194111

이름: 최민규

**CRISP-DM을 활용한 관심연구 초안 작성**

**주제 선정: 인스타그램 패션 인플루언서 포스팅이 실제 제품 구매에 미치는 영향**

|  |  |
| --- | --- |
| [1단계] 업무 이해 | 업무 목적: 인플루언서 마케팅의 효율성 향상, 제품 판매량 증대  상황 파악: 최근 제품 마케팅에서 SNS의 영향력이 증가하고 있는 상황. 이에 대한 이해와 분석이 필요하다는 판단.  데이터 마이닝 목표: 인플루언서 포스팅과 제품 구매 간 상관 관계 파악을 통해 제품 구매 유도 전략 도출.  프로젝트 계획: 인스타그램 API, 웹 크롤링 등을 사용 해 포스팅 데이터 수집, 머신러닝 모델을 활용 해 해당 포스팅과 제품 판매량 간의 영향 분석하고 이를 통해 결과 도출. |
| [2단계] 데이터의 이해 | 초기 데이터 수집: 인스타그램 API, 웹 크롤링을 통해 인플루언서의 포스팅 데이터 수집.  데이터 기술 분석: 수집된 데이터를 DB 형식으로 모으고, 이를 머신러닝 모델을 통해 분석.  데이터 탐색: 분석 과정 중 일정한 관계성이 존재하는지 확인하고 이를 탐색.  데이터 품질 확인: 분석을 통해 얻은 데이터에 특이점이 존재하는지 품질 확인 |
| [3단계] 데이터 준비 | 분석용 데이터 셋 선택: 인스타그램 패션 인플루언서의 포스팅 데이터 셋, 제품 구매 이력 데이터 셋을 선택.  데이터 정제: 머신러닝 모델 처리를 위해 해당 데이터셋들을 전처리.  분석용 데이터셋 편성: 각 데이터 셋에 분석에 필요한 각종 변수들 생성  데이터 통합: 인플루언서 포스팅 - 제품 구매 데이터 연결 및 통합.  데이터 포맷팅: 머신 러닝 모델 분석 활용을 위해 데이터를 적합한 형태로 인코딩하고 변환. |
| [4단계] 모델링 | 모델링 기법 선택: 머신 러닝으로 모델링 기법 선택  모델 테스트 계획 설계: 학습 데이터, 테스트 데이터 결정하고 분류.  모델 작성: 테스트 코드를 작성하고 적용  모델 평가: 작성한 테스트 코드의 실행 결과를 통해 모델의 성능 측정, 평가 |
| [5단계] 평가 | 분석 결과 평가: 주제로 선정했던 ‘인플루언서 포스팅 - 제품 구매에 미치는 영향’에 부합하는지 평가.  모델링 과정 평가: 모델의 출력을 보고 이용된 모델링 기법, 알고리즘을 평가하고 개선점 도출.  모델 적용성 평가: 개발된 모델이 실제로 적용될 수 있는지 평가. |
| [6단계] 전개 | 전개 계획 수립: 작성된 모델을 바탕으로 실제 인플루언서 마케팅 전략에 대한 구체적인 계획 수립  모니터링과 유지보수 계획 수립: 인플루언서 포스팅과 이에 대한 제품 구매에 대해 지속적으로 모니터링, 개선점을 찾아 유지보수 계획 수립.  프로젝트 종료 보고서 작성: 위의 평가 단계에서 도출된 전체적인 평가 내용과 개선점 등을 요약 해 프로젝트 종료 보고서 작성.  프로젝트 리뷰: 실제 인플루언서 마케팅에 미친 영향 등 좋은 점과 나쁜 점을 구별 해 리뷰. |