

개발 계획서

주제	주문에 따른 고객사 별 물량 수요 예측							
팀구성원	No.	구분	성명	소속	No.	구분	성명	소속
	1	팀장	홍승우	광주인공지능사관학교	2	팀원	박선희	광주인공지능사관학교
	3	팀원	정우철	광주인공지능사관학교	4	팀원	최해민	광주인공지능사관학교

1.1	배경 및 필요성
-----	----------

최근 이커머스 시장의 성장과 더불어 물류시스템을 최적화하여 배송시간을 단축시켜주는 e-풀필먼트 시스템이 주목받고 있습니다.

그래서 대규모 할인행사 등으로 대량의 수요가 발생하더라도 미리 효율적으로 대비할 수 있도록 솔루션이 필요하다고 생각했습니다.

1.2 목적

저희 팀은 평상시 고객사 별 물량의 흐름과 이벤트 발생 시 물량의 흐름을 미리 예측하고 대비함으로써 물류창고 재고관리를 더욱 효율적으로 운영할 수 있도록 하고, 궁극적으로 대량의 수요발생에도 변함없는 배송시간을 유지하는데 그 목적을 두고 진행했습니다.

1. 개 요

1.3 혁신성

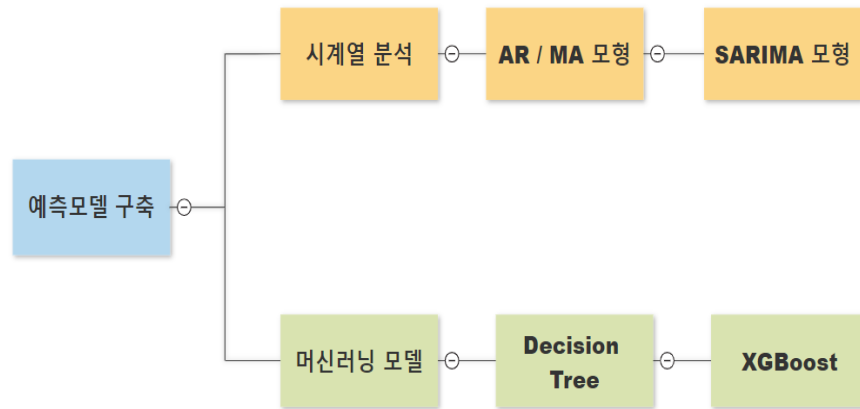
우선 저희 예측모델의 핵심은 머신러닝을 활용한 XGBoost 모델을 구축하면서 고객사별 최적의 모델을 생성하고 학습시켜 맞춤형 서비스가 가능하다는 점입니다.

또한, 저희가 개발한 솔루션은 ‘네이버 쇼핑 파트너 공식 블로그’ 크롤링을 통해 자동으로 이벤트 공지 게시물을 감지하여 이벤트 기간의 추출이 가능하기 때문에 추후에도 지속적인 운용이 가능하다는 장점을 가지고 있습니다.

그렇기 때문에 앞서 크롤링으로 확인된 해당 고객사의 이벤트 여부와 원
하는 예측 날짜를 입력하면 해당 날짜의 물량 수요가 결과값으로 산출되게
됩니다.

이러한 저희의 솔루션은 더욱 세밀한 고객사의 수요를 예측하면서 불필요한 비용과 시간을 단축시킬 수 있을 것입니다.

<p>2. 과제 수행</p>	<p>2.1 수행 방법</p> <p>1) 데이터 분석 및 전처리</p> <p>우선 저희 팀은 다 함께 모여 주어진 Data를 분석하고, 필요한 컬럼들을 분류하는 작업을 했습니다.</p> <p>그리고 고객사 코드(SHPR_CD)와 주문 날짜(BKG_DATE)를 기준으로 품목 수량(ITEM_QTY)을 모두 더하여 고객사별 물량 흐름을 파악할 수 있도록 Data를 가공했습니다.</p> <p>해당 물량과 수집한 외부 데이터(이벤트)의 상관관계를 분석했고, 각 팀원이 수행한 내용은 아래와 같습니다.</p> <ul style="list-style-type: none"> - 홍승우 : 공통된 분석을 위한 데이터 전처리 수행 - 박선흥 : 날짜, 가격, 품목순번, 요일, 네이버 스토어 이벤트 - 정우철 : KOSPI, KOSDAQ, 환율(\$) - 최해민 : 날짜, 네이버 데이터랩 (업종별 검색량) <p>이러한 과정에서 저희 팀은 CJ대한통운의 e-폴필먼트 사업에 있어 네이버가 가장 큰 협력관계라는 것을 알게 됐고, 관련 기사들을 통해 고객사 명단을 수집했습니다.</p> <p>그렇게 저희는 네이버 스마트스토어 이벤트, 대한통운 - 네이버 브랜드 스토어 관련 기사, 주어진 데이터 내 입력자ID (INS_ID)를 종합적으로 분석해 고객사를 특정할 수 있었습니다.</p> <p>또한, 이를 바탕으로 저희 팀은 고객사의 일일물동량과 네이버 스마트스토어 이벤트(레드위크, 브랜드데이, 쇼핑라이브 등)가 가장 상관관계가 높다고 판단했습니다.</p> <p>전체 Data에서 결측치를 제외하고, 물량 수요의 85% 이상을 차지하고 있는 상위 10개 기업을 위주로 이후 분석을 진행했습니다.</p>
-----------------	--



2) 모델 구축

이후 모델 구축에 있어 저희는 우선 종합적인 시계열 분석 및 AR, MA 시계열 예측을 위해 전체 Data를 세분화하여 상위 10개 기업에 대해 정상성 테스트를 진행했습니다.

그리고 해당 기업들을 대상으로 ACF, PACF를 진행하였으나 거의 모든 Lag에서 신뢰구간을 벗어나지 못하여 종합적으로 유의미한 결과값을 얻을 수 없었습니다.

이후 이벤트 물량에 Winsorize한 Data로 추가 분석을 진행한 결과, 평균과 분산이 이전 Data에 비해 시간에 따라 비교적 일정한 모습을 보이고 있으며, 잔차가 전보다 정규분포를 띄고 있어 SARIMAX 모델로 시계열 분석이 가능할 것이라고 생각했습니다.

SARIMAX 모델을 활용한 시계열 분석을 진행하였으나, 상위 10개 고객사 중 8개의 경우 P-value가 유의수준인 0.05%를 통과하지 못하여 시계열 모델이 부적합하다고 판단했습니다.

마지막으로 여러 개의 Decision Tree를 조합하여 앙상블한 알고리즘인 XGBoost를 활용하여 모델 구축을 진행했습니다.

이벤트라는 설명변수를 적절히 활용하고 파생변수를 만들어 점차 모델을 최적화했습니다. 그 결과 MAPE 기준 약 19.48이 나왔으며 이벤트라는 outlier이 있었음에도 불구하고 훌륭한 성능을 보여줬습니다.

이 과정에서 고객사별로 학습을 시켰으며 이벤트 값을 넣을 수 있도록 하여 사전에 크롤링한 자료를 활용해 더욱 세밀한 결과값을 도출할 수 있도록 최종적으로 모델을 구축했습니다.

	<p>위 과정에서 팀원이 수행한 내용은 아래와 같습니다.</p> <ul style="list-style-type: none"> - 홍승우 : 시계열 분석 평가 및 검증, 이벤트 데이터 수집 - 박선훈 : 시계열 분석(AR, MA 모형), XGBoost 모델 구축 - 정우철 : 이벤트 크롤링, SARIMAX 모델 구축, 개발계획서 작성 - 최해민 : 네이버 쇼핑블로그 이벤트 크롤링, 데이터 시각화 작업
3. 활용방안 및 기대효과	<p>3.1 활용방안</p> <p>이번에 저희가 만든 솔루션은 고객사별로 보다 정밀한 이벤트 물량 예측이 가능하기 때문에 미리 필요한 물량만큼 예비 재고를 준비하여 주문 즉시 운송할 수 있습니다.</p> <p>3.2 기대효과</p> <p>기존 CJ대한통운의 e-풀필먼트 솔루션과 이벤트에 특화된 저희의 고객사 맞춤 예측 솔루션을 활용한다면 더욱 섬세하고 효율적인 물류시스템 운용이 가능할 것이라고 생각합니다.</p> <p>효율적 물류시스템 운용을 통해 불필요한 인력, 장비, 운송 비용의 지출을 줄일 수 있으며, ‘오늘 출발’이라는 고객과의 약속을 365일 지킬 수 있을 것입니다.</p>