

온프레미스 환경 기반 MLOps 파이프라인을 활용한 아파트 시세 예측 및 자동화

김수현, 이남경, 박현석, 김지현, 김희원, 이건희*
에이치디씨랩스

{soohyun.kim, nk_lee, hyunskki, jh991219, ive2go, Gunhee_Lee}@hdc-labs.com

MLOps Pipeline for Apartment Price Prediction in an On-Premises Environment

Soohyun Kim, Namkyeong LEE, Hyunseok Park, Jihyeon Kim, Heewon Kim,
Gunhee Lee*
HDC LABS

요 약

본 연구는 온프레미스 환경에서 아파트 시세 예측을 위한 MLOps 파이프라인을 설계·구현하고 검증하였다. 주 단위로 수집되는 데이터는 회귀 기반 보간과 이상치 탐지·제거 과정을 통해 품질을 보완하였다. 파이프라인은 쿠버네티스 기반 자동화를 활용하여 데이터 준비, TimeLLM 모델의 미세조정, 모델 평가, MLflow 기반 기록 및 버전 관리, 배포까지 엔드투엔드 프로세스를 구현하였다. 매주 성능이 우수한 모델을 자동 선별하여 시장 변동에 대응하고, 주 1회 배치 학습을 루프를 통해 온프레미스 환경에서도 지속적 학습·배포·모니터링이 가능케 한다. 이를 통해 본 연구는 실무 지향적 참조 아키텍처를 제안한다.

I. 서 론

아파트 시세 예측은 정책 수립과 투자 의사결정에 필수적인 과제로, 머신러닝 및 딥러닝 연구가 활발하다.[1][2] 그러나 기존 연구는 정적인 데이터셋을 중심으로 성능 평가에 집중하는 경향이 있으며, 온프레미스 환경에서 지속적 학습과 자동 배포가 가능한 MLOps 적용 사례는 상대적으로 부족하다.

아파트 가격은 정부 정책, 금융 규제, 금리 변동 등 외부 요인에 의해 변동성이 존재하므로, 모델을 주기적으로 업데이트하지 않으면 예측 정확도가 저하된다. Hahn et al.[3]는 부동산 시장이 정책 변화, 재난, 거시경제 불확실성 등 외부 충격에 의해 가격과 거래량의 변동성이 뚜렷하게 확대됨을 보여준다. 이러한 충격은 즉각적이거나 점진적으로 발생하며 다양한 경로를 통해 누적되기 때문에, 정적으로 학습된 단일 모델 접근으로는 대응에 한계가 있다. 실제로 코로나 팬데믹, 급격한 금리 인상 등과 같은 사건에서는 정적 모델의 성능 저하 경향이 관찰된다.

부동산 거래 데이터는 개인정보와 민감 정보가 포함되어 있어 외부 클라우드로의 이전이 어렵고, 내부망에서 안전하게 운영되어야 한다. 또한 온프레미스 환경은 네트워크 분리와 한정된 연산 자원이라는 제약 속에서 안정성과 효율성을 동시에 확보해야 하는 특수성을 가진다. 따라서 단순 예측 모델을 넘어 외부 충격을 지속적으로 모니터링하고 신속히 반영할 수 있는 동시에, 데이터 보안과 내부망 운영 요건을 충족하는 온프레미스 기반 MLOps 체계가 필요하다.

이에 본 연구는 온프레미스 환경에서 TimeLLM[4] 기반 시계열 예측 모델의 주간 미세조정과 쿠버네티스

및 MLflow 를 활용한 자동화 MLOps 파이프라인을 구축·운영함으로써, 변화 대응력을 강화하고 지속적 학습·배포가 가능한 아키텍처를 제시한다.

II. 본론

1. 데이터 수집 및 전처리

연구에 활용된 데이터는 국내 부동산 데이터 플랫폼을 통해 주기적으로 수집되었으며, 부동산 거래 관련 다양한 속성을 포함한다. 데이터 품질 향상을 위해 결측치는 회귀 기반 보간법으로 보완하였고[5], 이상치는 규칙 기반 탐지 방법으로 제거하였다. 대용량 시계열 데이터 처리에는 PySpark 기반 분산 전처리를 적용하였으며, 분석 대상은 최근 5 년간 누적 데이터를 기본으로 하되 매주 새롭게 수집되는 자료가 반영되도록 지속적으로 업데이트된다.

2. MLOps 파이프라인 설계 및 모델 관리

본 시스템은 주간 단위로 갱신되는 부동산 데이터를 입력받아, 전처리, 보간, 데이터 분리, TimeLLM 학습과 평가로 이어지는 MLOps 파이프라인으로 구현되었다. 전 과정은 컨테이너화되어 재현성과 확장성을 확보하며, 모든 실험과 모델 산출물은 MLflow 에서 통합적으로 관리된다.

모델 및 데이터 버전, 하이퍼파라미터, 성능 지표(MAPE, MSE, RMSE)는 체계적으로 기록된다. 학습 완료 후 새 모델과 기존 운영 모델의 성능을 비교해 기존 모델이 더 우수할 경우 운영 모델로 자동 승격하고,

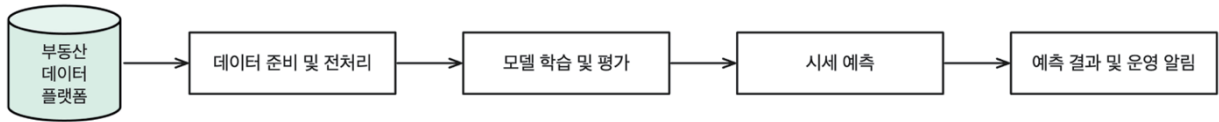


그림 1. 아파트 시세 예측을 위한 MLOps 파이프라인 및 운영 알림 체계 개요

기존 모델은 아카이브 처리한다. 주간 아파트 시세 예측 및 배포는 오직 운영 모델을 통해 수행된다.

제안한 파이프라인의 전체 구조는 그림 1 과 같이 데이터 준비 및 전처리, 모델 학습·평가, 시세 예측, 그리고 후술할 예측 결과 전달과 운영 알림 단계로 구성된다.

모델 관리와 배포 과정의 신뢰성을 높이기 위해, 학습·평가·예측 단계별로 사용되는 데이터의 시점을 명확히 구분하였다. 구체적인 데이터 시점 전략은 다음과 같다. 주차를 W_t 라 할 때, 학습에는 W_{t-2} 까지의 데이터를 사용하고, 평가는 W_{t-1} 데이터를 활용한다. 예측은 금주 입력 데이터 X_t 를 기반으로 아파트 시세 \hat{Y}^t 를 산출한다. 이러한 과정을 수 식으로 표현하면 다음과 같다.

$$Model_t = f(Data \leq W_{t-2})$$

$$Eval(Model_t, W_{t-1})$$

$$\hat{Y}^t = Model_t(X_t)$$

시점 분리 전략은 최신 데이터를 반영하면서도 모델 안정성을 확보하고, 시장 변동에 따른 성능 저하를 최소화한다. 또한 주간 배치 학습을 통해 우수한 성능을 보인 모델만을 자동으로 선별·배포하여 불필요한 모델 교체를 방지함으로써 안정적인 서비스 운영이 가능하다.

모델 성능 유지와 더불어, 온프레미스 환경에서는 보안 측면에서의 데이터 및 모델 자산 보호가 핵심 과제로 고려되어야 한다. 이에 따라 데이터와 모델 아티팩트는 내부망 전용 저장소에 안전하게 보관되며, 모델 관리 및 실험 기록 시스템은 외부 네트워크와 완전히 격리되어 있다. 또한 인증 및 권한 관리 체계를 기반으로 접근이 통제되어 온프레미스 환경의 보안 요구사항을 충족한다.

3. TimeLLM 기반 시계열 예측

본 연구는 대규모 언어 모델을 시계열 데이터 특성에 맞게 최적화한 TimeLLM 을 도입하여, 주간 단위로 수집된 데이터를 반영해 미세조정을 수행한다. 이를 통해 최신 시장 트렌드를 효과적으로 반영하며, 주 단위 데이터 특성에 적합한 예측 정확도를 향상시킨다.

4. 운영 안정성 및 자동 알림 체계

앞서 제시한 파이프라인의 마지막 단계인 ‘예측 결과 전달 및 운영 알림’은 단순한 예측 수행을 넘어 실제 운영 안정성과 직결된다. 본 연구의 파이프라인은 온프레미스 환경에서 주간 단위 배치 학습·배포를 자동화함으로써 최신 시장 변동을 신속히 반영한다. 파이프라인의 안정적 운영을 위해 핵심 성능 지표(MAPE, MSE, RMSE)를 지속적으로 모니터링하며, 모델 성능 저하나 이상 징후 발생 시 신속 대응을 가능하게 한다.

특히, 파이프라인 수행 성공 및 실패 시 이메일 및 메신저 자동 알림 기능을 구현하여 운영 담당자가 즉각적으로 상태를 파악하고 조치할 수 있도록 하였다. 이러한 자동 알림 체계는 온프레미스 환경의 모니터링 인프라 제약을 보완하며, 장애 조기 탐지와 복구 시간을 단축하는 효과적인 수단으로 작용한다.

이와 같은 운영 안정성 확보와 자동화된 알림 시스템 결합은, 온프레미스 환경에서도 신뢰할 수 있는 MLOps 파이프라인 운영을 가능케 하여 실무 적용성을 높인다.

III. 결론

본 연구는 온프레미스 환경에서 부동산 시세 예측에 특화된 MLOps 파이프라인 구축 및 운영을 실증하였다. 체계적인 데이터 전처리와 TimeLLM 기반 시계열 예측 모델의 주기적 미세조정, 쿠버네티스 및 MLflow 활용 자동화는 최신 시장 변동에 실시간 근접 대응을 가능케 했다. 본 시스템은 주 1 회 배치 학습과 실시간 성능 모니터링을 통해 안정성과 지속 가능성을 확보, 실무 적용 가능한 참조 아키텍처를 제공한다.

나아가 본 연구는 온프레미스 환경의 보안성과 안정성을 기반으로 하여, 외부 클라우드에 의존하지 않고 민감한 부동산 데이터를 안전하게 관리·분석할 수 있는 체계를 제시하였다. 이는 데이터 주권과 내부망 운영이 필수적인 환경에서도 활용 가능한 확장성을 보여준다. 향후 외부 변수 통합이나 모델 확장을 통해 예측 성능과 활용성을 더욱 고도화할 수 있다.

참 고 문 헌

- [1] H. Kim, H. Yoo, and H. Oh, "Apartment price prediction using deep learning and machine learning," Journal of Information Processing Systems, vol. 12, no. 2, pp. 59–76, Feb. 2023.
- [2] S.-H. Nam, T.-H. Han, I.-J. Kim, and E.-J. Lee, "Real-estate price prediction in South Korea via machine learning modeling," Journal of the Korea Institute of Internet, Broadcasting and Communication, vol. 20, no. 6, pp. 15–20, 2020.
- [3] A. Hahn, S. Omerovic, and S. Wlatl, "A framework to monitor the effects of external shocks on housing markets," arXiv preprint, arXiv:2502.03012 [econ.GN], Feb. 2025.
- [4] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in Proc. 12th Int. Conf. on Learning Representations (ICLR), Apr. 2024.
- [5] D. Lee, S. Lee, H. Park, H. Kim, and G. Lee, "T-Forest: Time series forecasting of real estate prices with complementary missing value imputation," in Proc. Korean DataBase Conf. (KDBC), 2024, pp. 77–86.