T4056_김찬호: 개인 회고

- **나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?**
  - **우리 팀과 나의 학습목표는 무엇이었나?**

  첫 팀 프로젝트인 만큼, 팀 프로젝트의 여러 과정을 경험하면서 익숙해지고 프로젝트에 필요한 여러 역량을 기르는 것을 목표로 삼았습니다.

  - **개인학습측면**

  가장 먼저 베이스라인 코드를 이해하고 데이터의 특성을 고려해 EDA를 진행했습니다. 그 후에비정형 데이터를 활용하는 딥러닝 모델들과 베이스라인 외 모델들을 탐색했습니다.

  모델의 성능을 향상시키기 위해 모델의 특성에 맞추어 데이터를 전처리하고, Optuna를 활용하여 하이퍼 파라미터를 최적화했습니다. 추가로 교차 검증을 적용하여 여러 개의 모델을 생성하고 서로 다른 데이터 셋에 대해 학습시킨 뒤 결과를 앙상블하여 모델을 고도화 하기 위해 노력했습니다.

  - **공동학습측면**

  Github를 활용하여 서로의 코드를 리뷰하고 결과물을 공유했습니다. 또, 공유하고 싶은 내용을 슬랙을 통해 바로바로 전달하고 상호간에 도움을 주었습니다.

  프로젝트 기간 동안 매일매일 그 날의 진행계획과 성과 등을 노션에 정리하고 공유했습니다.

- **나는 어떤 방식으로 모델을 개선했는가?**
  - **사용한 지식과 기술**

  프로젝트 기간 중 제가 담당하고 시간을 투자한 모델은 TabNet입니다.

  부스팅 기반 모델과 딥러닝의 장점을 아우르는 특성을 가지고 있다는 점이 매력적이었고, 본 프로젝트에서 뛰어난 성능을 얻을 수 있는 모델이라고 생각했습니다.

  Scikit-learn 기반의 모듈을 사용하여 모델을 구현했고, 범주형 변수를 따로 입력받아 처리하는 TabNet의 특성에 맞게 데이터를 전처리하고 입력했습니다.

  하이퍼 파라미터를 최적화하기 위해 튜닝 tool인 Optuna를 사용했고, 교차 검증 기법을 통해 여러 조합의 train/valid 데이터 셋에 대해 모델을 학습시키고 결과물을 앙상블해서 더 나은 성능을 얻고자 했습니다.

- **내가 한 행동의 결과로 어떤 지점을 달성하고, 어떠한 깨달음을 얻었는가?**

데이터와 모델의 특성에 따라 같은 모델과 데이터라도 어떻게 처리하고 조정해주냐에 따라서 결과가 크게 달라질 수 있고, 특정 모델과 데이터의 조합은 서로 특성이 맞지 않아 좋은 성능을 얻기 어려울 수 있다는 것을 알게되었습니다.

구조가 단순해 보이는 기본 모델이라도 데이터를 알맞게 전처리하고 모델을 조정하면 만족스러운 성능을 보여줄 수 있지만, 비교적 최근에 제안된 고도의 모델이라도 적합한 데이터에 사용되지 않는다면 좋은 성능을 보여주는 것은 어려울 수 있다는 점을 알게되었습니다.

- **전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?**

이전까지는 이러한 대회형 프로젝트의 경험이 없었고, 주어진 문제를 어떻게 해결해야 할지 막연함이 있었습니다. 작은 목표부터 설정하고 성취해나가면서 자신감이 생겼고 어느 순간 내가 하고 있는 일이 팀에 기여하고 있다는 느낌이 들었습니다.

- **마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?**

TabNet은 딥러닝 기반 모델이지만, 모듈을 통해 구현했을 때 모델에서 조정할 수 있는 부분에는 명확한 한계가 있습니다. 다양한 비정형 데이터나 시퀀싱 데이터를 입력으로 넣어주거나 end-to-end로 학습시켜 입력데이터 처리도 한 번에 진행하고, 모델 내부 구조를 수정하는 등 딥러닝의 큰 장점들을 제대로 활용하지 못한 것 같아 아쉬움이 남습니다.

또한 TabNet을 다루는 과정에서 모델과 데이터 셋에 대한 이해가 크지 않아 해당 모델이 프로젝트에 효과적인지 판단하기가 어려웠고, 적절하게 데이터를 처리하고 모델을 조정해서 프로젝트에 최적화된 지점까지 도달하는 일이 순조롭지 않았다.

- **한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해볼 것은 무엇일까?**

간단한 모델이라도 모듈에 의지하지 않고 바닥부터 구현해서 문제 해결에 적용해보는 시도를 하고 싶다. 또한 특히 딥러닝 모델의 경우 비정형 데이터를 활용하고 부족한 점을 보완하는 등 다양한 내부 구조 조정을 시도해보고 싶다.

그리고 이번 프로젝트를 통해 얻은 인사이트로부터 주어진 문제를 해결하기 위해 데이터를 어떻게 활용하고 어떤 부분에 대한 모델의 표현력이 필요한지, 한 모델을 보완하기 위해 어떻게 접근해야할지를 논리적으로 접근하여 프로젝트 해결 과정을 논리적인 바탕 위에 탄탄하게 진행하고 싶다.

또 모델을 사용할지에 대한 판단을 내릴 때, 특정 하이퍼 파라미터에 민감한 모델일 수 있으므로 해당 모델에 맞춘 데이터 전처리 방법과 하이퍼 파라미터 튜닝까지는 진행해봐야겠다는 생각이 들었다.