**통계학 기반 금융 데이터 분석 프로젝트 기획안**

|  |  |
| --- | --- |
| 팀명 | 사랑합니다 |
| 주제명 | 딥러닝 기반 가상화폐 자동매매 시스템 |

1. **역할분담**

|  |  |
| --- | --- |
| 이름 | 역할 |
| 박훈석 | 팀장 / 프로젝트 관리(PM) / 도메인 분석 |
| 이현 | 데이터베이스 설계 및 관리 / 딥러닝 모델링 |
| 정민관 | 데이터베이스 설계 및 관리 / 업비트 API 연동 및 관리 / 딥러닝 모델링 |
| 신영섭 | 데이터 분석 및 전처리 / 백테스팅 시스템 구현 |

**2. 주제 선정 배경**

최근 암호화폐 시장은 급격한 성장과 함께 높은 변동성이라는 특징을 보이며 투자자들의 큰 관심을 받고 있습니다. 특히 비트코인은 최초의 암호화폐로서 시장에 큰 영향력을 행사하지만, 예측하기 어려운 가격 변동은 투자 결정에 어려움을 야기합니다. 따라서, 시장 참여자들의 불확실성을 줄이고 합리적인 의사결정을 지원하기 위해 신뢰성 높은 비트코인 가격 예측 모델에 대한 필요성이 강조되고 있습니다.

딥러닝 기술이 발전했음에도 불구하고, API를 통해 쉽게 얻을 수 있는 시장 데이터를 활용한 심층 학습 모델 연구는 아직 미흡한 수준입니다. 또한, 다양한 피처 선택 방법이 예측 모델 성능에 미치는 영향에 대한 실증적 분석 역시 부족한 상황입니다.

이에 본 프로젝트는 딥러닝 기술의 잠재력을 최대한 활용하여 비트코인 가격 방향 예측 모델의 정확도를 높이는 것을 목표로 합니다. 다양한 거래소 또는 데이터 제공 플랫폼에서 API를 통해 제공되는 가격 데이터, 거래량, 기술 지표 등의 시장 데이터를 활용하여 기존 연구의 한계를 극복하고자 합니다. 더불어, 다양한 피처 선택 방법을 적용하여 불필요한 변수를 제거하고 모델의 일반화 성능을 향상시키는 데 집중할 것입니다. 궁극적으로 본 프로젝트는 실제 투자 환경에 적용 가능한 효율적인 예측 모델을 개발하고, 이를 기반으로 한 거래 전략의 수익성을 평가하여 실질적인 투자 가치를 창출하는 것을 목표로 합니다.

**3. 프로젝트 목표**

1. **모델 성능 비교:** CNN-LSTM, LSTNet, TCN 등 다양한 딥러닝 모델의 성능을 비교 분석하여 최적의 모델을 선정합니다.
2. **피처 선택 방법 효과 검증:** Boruta, 유전 알고리즘(GA), LightGBM 등 다양한 피처 선택 방법이 예측 모델 성능에 미치는 영향을 실증적으로 분석합니다.
3. **거래 전략 수익성 평가:** 개발된 예측 모델을 기반으로 한 다양한 거래 전략(매수 및 매도, 가격 보호, 롱앤숏 등)을 백테스팅하여 실제 시장에서의 수익성을 평가합니다.
4. **연구 결과의 실용화:** 개발된 모델과 거래 전략을 바탕으로 실제 투자에 적용 가능한 가이드라인을 제시합니다.

**4. 활용 데이터**

* 업비트 API

**5. 분석 방안**

**5.1. 데이터 수집 및 전처리**

* **데이터 소스:** 업비트 API를 통해 비트코인 가격, 거래량, 기술 지표 데이터 수집
* **데이터 전처리:**
  + 누락값 처리: MCAR (Listwise 삭제), MNAR (Linear Imputation)
  + 데이터 분할: 학습 데이터 (80%), 테스트 데이터 (20%) (시간 순서 유지)
  + 피처 표준화: 입력 피처의 스케일 조정
  + 이진 분류 변환: 가격 상승 (1), 가격 하락/유지 (0)
  + 데이터 증강 고려: 필요 시 데이터 증강 기법 도입

**5.2. 피처 엔지니어링 및 선택**

* **피처 생성:** 업비트 API 데이터 기반 다양한 파생 피처 생성 (예: 이동평균, RSI, MACD 등)
* **피처 선택:**
  + Boruta: 랜덤 포레스트 기반 래퍼 피처 선택 방법
  + GA: 유전 알고리즘 기반 래퍼 피처 선택 방법
  + LightGBM: 그래디언트 부스팅 기반 임베디드 피처 선택 방법
* **피처 중요도 분석:** 각 피처의 예측 모델 성능 기여도 분석

**5.3. 딥러닝 모델 개발 및 학습**

* **모델 선정:** CNN-LSTM, LSTNet, TCN 등 시계열 데이터 분석에 적합한 딥러닝 모델 사용
* **모델 구조 설계:** 각 모델의 최적 구조 (레이어 수, 노드 수, 활성화 함수 등) 설계
* **하이퍼파라미터 튜닝:** 임의 탐색 방법 등 다양한 하이퍼파라미터 튜닝 기법 활용
* **모델 학습:** 학습 데이터셋을 사용하여 모델 학습
* **교차 검증:** 교차 검증을 통해 모델의 일반화 성능 평가

**5.4. 모델 평가**

* **평가 지표:** 정확도 (Accuracy), 정밀도 (Precision), 재현율 (Recall), F1 점수, ROC AUC, Matthews 상관 계수 (MCC) 등 다양한 평가 지표 사용
* **모델 성능 비교:** 학습된 모델의 성능을 비교 분석하여 최적 모델 선정
* **통계적 검정:** Wilcoxon signed-rank 및 Friedman 검정 등을 활용하여 모델 간 유의미한 성능 차이 검증

**5.5. 백테스팅 및 거래 전략 평가**

* **거래 전략 구현:**
  + 매수 및 매도 (Buy and Sell) 전략
  + 가격 보호 기능이 있는 매수 및 매도 전략
  + 롱앤숏 전략
  + 기준 전략 (MACD)
* **백테스팅:** 과거 데이터를 사용하여 거래 전략의 성능 시뮬레이션
* **거래 시뮬레이션:** 테스트 데이터를 사용하여 거래 시뮬레이션 및 수익률, 샤프 비율, 최대 하락률 (MDD) 등 다양한 지표 분석
* **수익성 평가:** 다양한 거래 전략의 수익성 비교 분석

**6. 결론**

본 프로젝트를 통해 우리는 딥러닝 기술의 가상화폐 시장 적용 가능성을 실증적으로 검증하고자 합니다. 딥러닝 기술을 통해 투자자에게 가상화폐 시장의 방향성을 예측하고, 이를 기반으로 객관적인 투자 전략 수립이 가능할 것으로 기대합니다. 특히 우리는 이번 프로젝트를 통해 인간의 비이성적인 투자 성향에서 벗어나 데이터 기반의 객관적인 분석과 투자 전략을 연구하고, 실제 가상화폐 시장의 데이터를 다루는 과정에서 얻게 될 개발 경험이 향후 실무에 중요한 자산이 될 것으로 기대합니다.