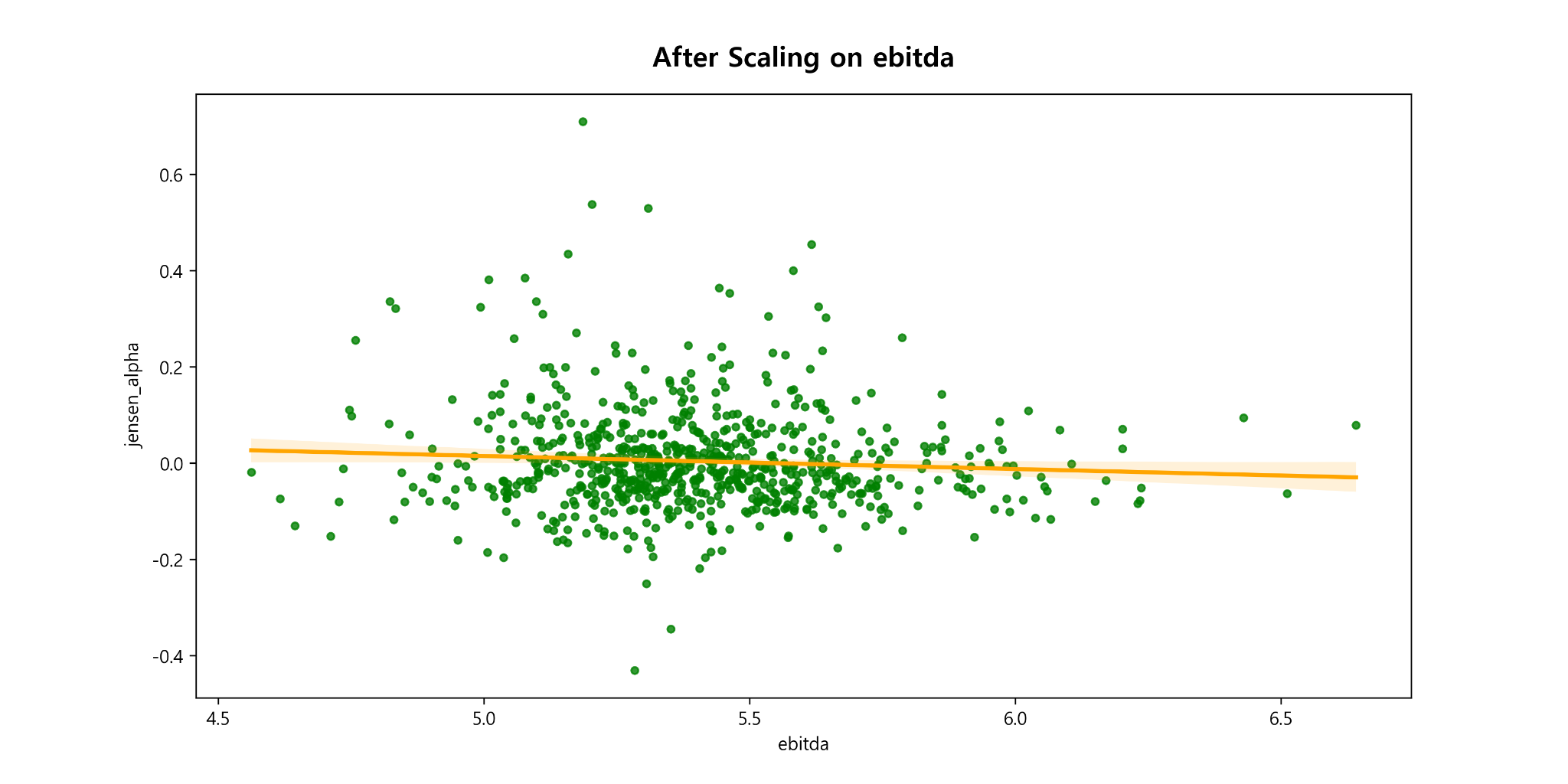
**주제 변경사유**: 기존에 데이터를 활용하여 주가 예측 및 우상향 기업을 주식 입문자들에게 제공 하는 서비스를 만들고자 하는 취지에서 시작하였습니다.

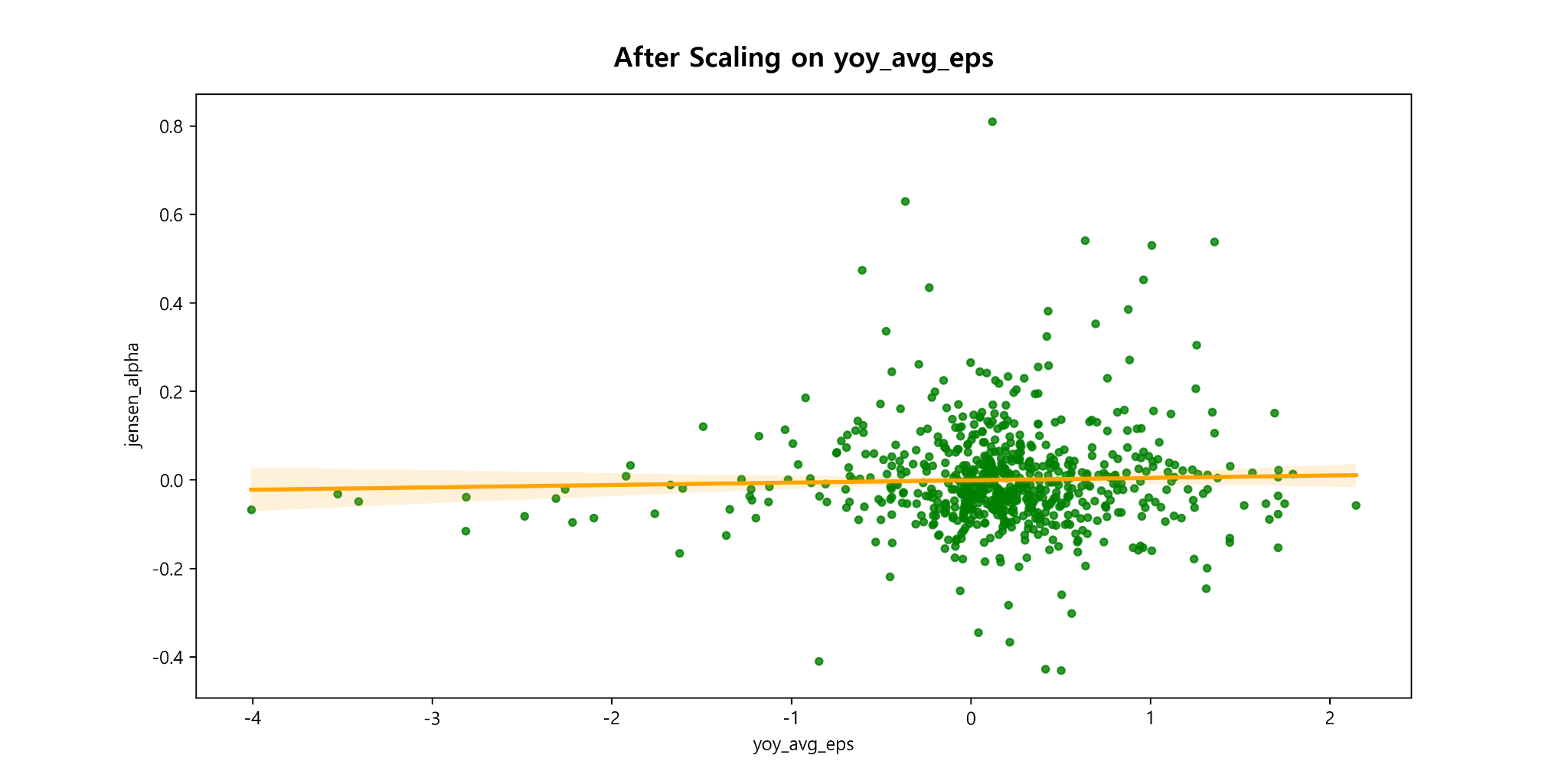
그러나 재무적 지표(feature)와 젠센알파(target)를 시각화 한 결과 및 선형 회귀를 통한 계수 유의성 검증결과를 결합하여 feature 영향을 분석해 보니 대부분의 feature 가 영향력이 없다는 것을 발견하였습니다.

(train feature 기간 2012 ~ 2018 연간 평균 성장률 및 최근 3년 재무데이터,

train target 기간 2019년 분기 평균 젠센 알파)

기존 feature 에 비정형 데이터(뉴스 기사 등)을 결합하여 성능을 올려보고자 하였으나, 재무데이터가 1년 단위의 데이터인 점을 고려할 때 비정형 데이터 기간과 호환이 되지 않는다는 문제를 발견하였습니다. 이에 저희 팀은 기존 코드를 활용하여 주가 관련 프로젝트를 할 수 있는 데이터 셋을 찾다가 데이콘 공모전 데이터 셋을 사용하기로 했습니다.

<그림 : Robust scaling 후 일부 feature plotting>



**<인공지능 기반의 패턴 분석을 통한 주가 예측>**

**1. 질문 진행에 앞서 프로젝트에 대한 설명/발표 부탁 드립니다.**

**- 어떤 문제를 풀려고 하는지? 목표**

(문제): 정형데이터에만 의존했던 주가 예측 모델의 한계점을 비정형데이터를 활용해 보완

(목표): 기존 주가데이터 예측에 자연어 처리 및 감성분석 피쳐를 추가해 예측 정확도를 개선

**- 어떤 데이터를 사용하였는지?**

정형데이터 : 종목 주가 및 거래량 데이터 (데이터 공모전측 제공)

<https://dacon.io/competitions/official/235800/data>

비정형데이터 : 텍스트 데이터(뉴스 기사 제목 및 증권사 레포트 자료)

**- 어떤 Flow로 데이터를 전처리 하였는지? 중간 데이터 형태도 함께 설명**

진행 중 입니다

**- 학습은 어떤 방식으로 진행하고자 하는지?**

1. 부스팅 모델
2. LSTM 모델
3. MLP 모델 (개발한 모델 + Kaggle 참조 모델)
4. 앙상블 모델

**2. 추가적인 궁금증**

**- 과제외 Data Engineering 관련 내용**

정형데이터 처리 시, NA값 처리 및 이상치에 대한 처리가 어렵다고 느껴집니다. 이런 과정에서 싸이킷런 또는 타 모듈의 알고리즘을 적극 활용하시는지요? 아니면 시각화 후 개별 샘플을 직접보고 판단하시는지요?

**- 현업 관련 질문**

(1) 현업에서 현재 주목하고 있는 기술로 배우기를 추천하는 것이 있는지요?(ex. computer vision, NLP)

(2) 신입채용 시 대학원 석사이상이 거의 필수적으로 요구되는지요?

(3) 특정 산업에 따라 요구되는 역량이 다른데, (예로 금융에서는 시계열이나 정형 데이터 처리 역량이 더욱 요구되는 것으로 들었습니다) 신입 채용 시에는 모든 분석 역량을 갖춘 인재를 우선 선발하는 쪽으로 고려하시는지요?

**- 그외 개발과 관련 내용**

LG 연구소 이효성 연구원

pencial@naver.com