By Stephan O. A., Mojtaba P., Julia M., Roko R. R.

Urban Air Pollution Challenge

Project - World Air Quality Index

source: https://zindi.africa/competitions/zindiweekendz-learning-urban-air-pollution-challenge

- The **World Air Quality Index** project is a non-profit project started in 2007. Its mission is to promote air pollution awareness for citizens and provide a unified and world-wide air quality information. The project is providing transparent air quality information for more than 100 countries, covering more than 12,000 stations in 1000 major cities, via those two websites: aqicn.org and waqi.info

Task- Hypothesis

Fact

- Recent news articles stating that air quality has improved due to COVID-19 (world wild, because ppl staying at home)

Contradiction

- In many African cities, air quality seems to be getting worse (while ppl are staying at home)

Hypothesis

- The air quality is multifactorial and will depend on local weather phenomena such as windspeed, solar intensity and chemical composition of the air.

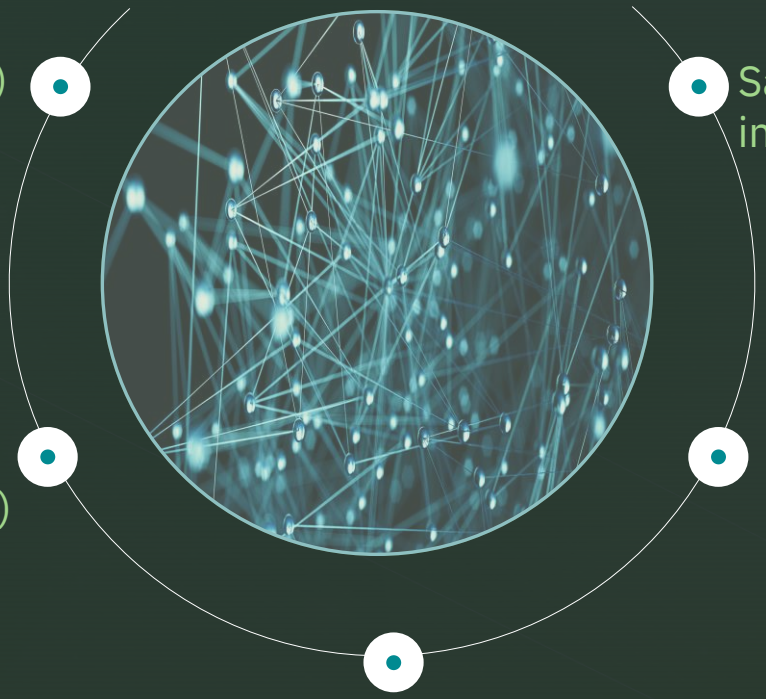
Task- Goal predict PM2.5

- Most studies indicate PM2.5 at or below 12 µg/m3 is considered healthy with little to no risk from exposure. If the level goes to or above 35 µg/m3 during a 24-hour period, the air is considered unhealthy and can cause issues for people with existing breathing issues such as asthma.
- Goal is to use the satellite information to predict PM2.5 particulate matter concentration (a common measure of air quality that normally requires ground-based sensors to measure)

Data - Description

data covers three months (01-03)

Satellite data from various pollutants in the atmosphere:
Key measurements like NO2 in different atmospheric hights



Ground-based air quality sensors measure the target variable (y actual)

Metadata like the satellite altitude excluded

Global Forecast System (GFS) for weather data
Humidity, Temperature, Wind Speed,
which are used as inputs

Data - Description

Model – Evaluations (RMSE)

- The error metric for this competition is the Root Mean Squared Error (RMSE)
- The result should be a combination of space, time and PM2.5 concentration

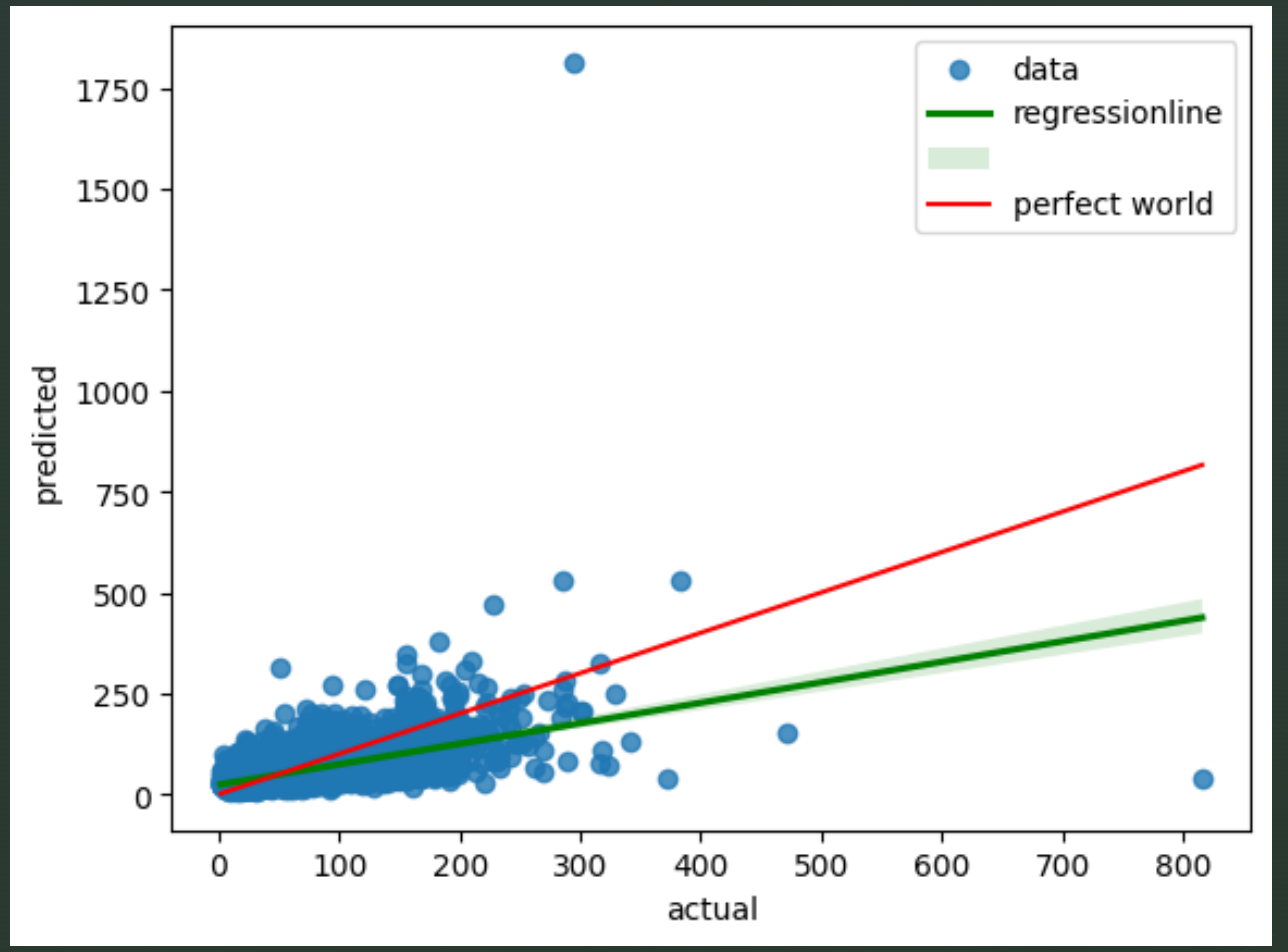
Place_ID X Date	target
00S9LVX X 2020-01-02	2
00S9LVX X 2020-01-03	91
00S9LVX X 2020-01-04	34

Model – Introduction

- Base model: Linear regression (free)
- Advanced base model: Elestic Net regression (penalty for overfitting)
- Simple model: Singuar Decision Tree
- Ensamble model: Random Forest
- Advanced ensamble model: Stacking

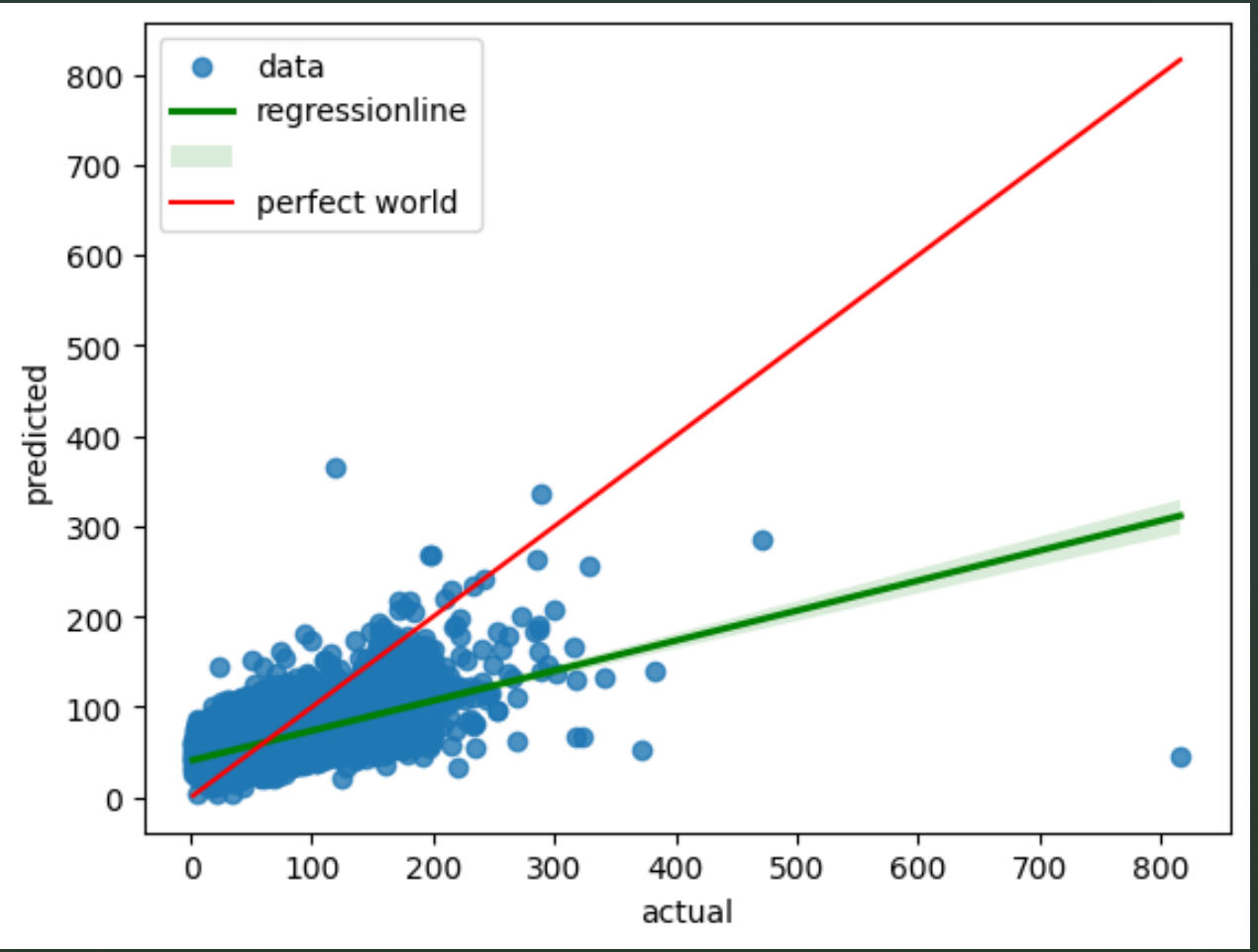
Base Model: Linear regression

(not penalty)

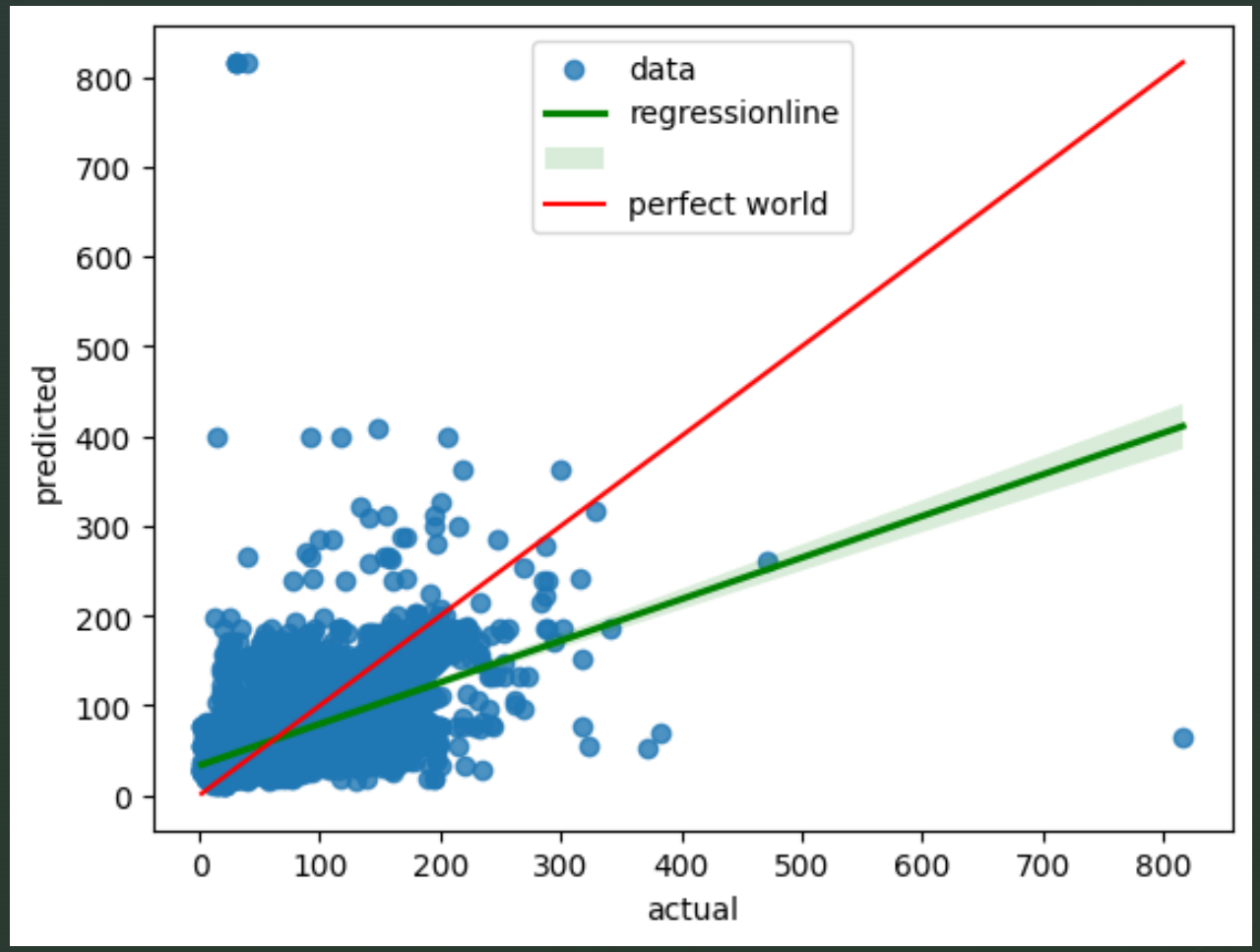


Advanced Base Model: Elestic Net Regression

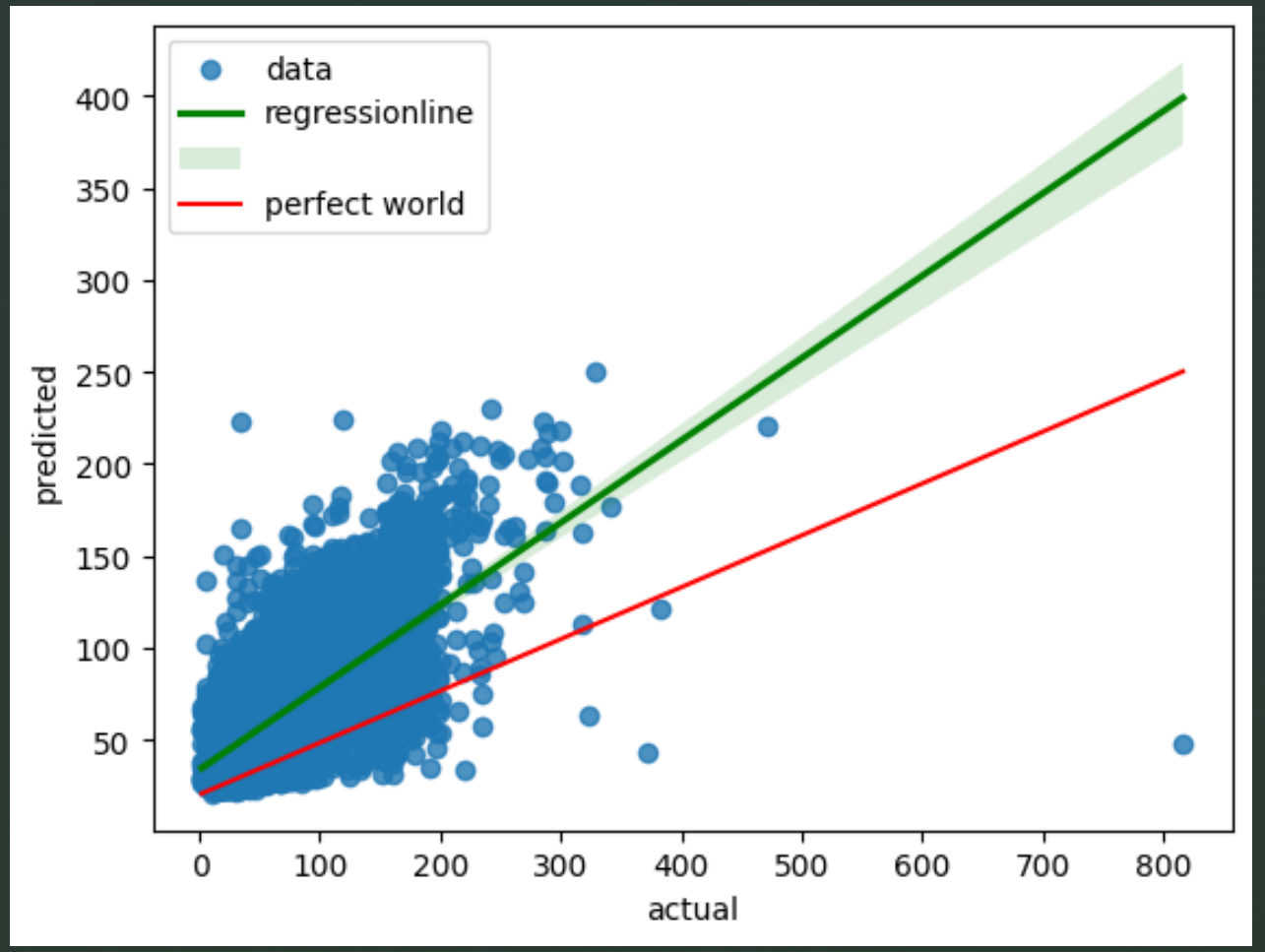
(penalty for overfitting)



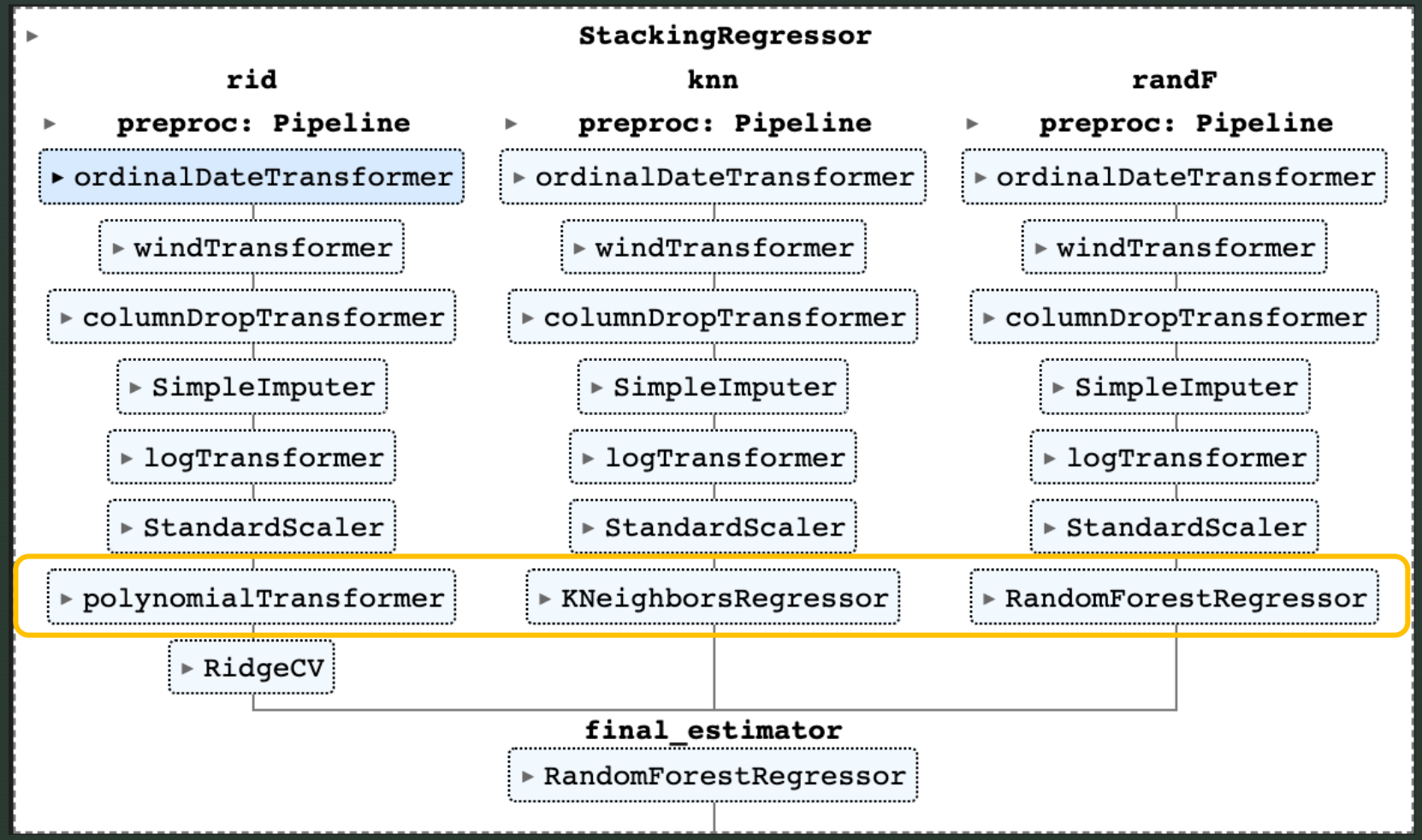
Simple Model: Singular Decision Tree



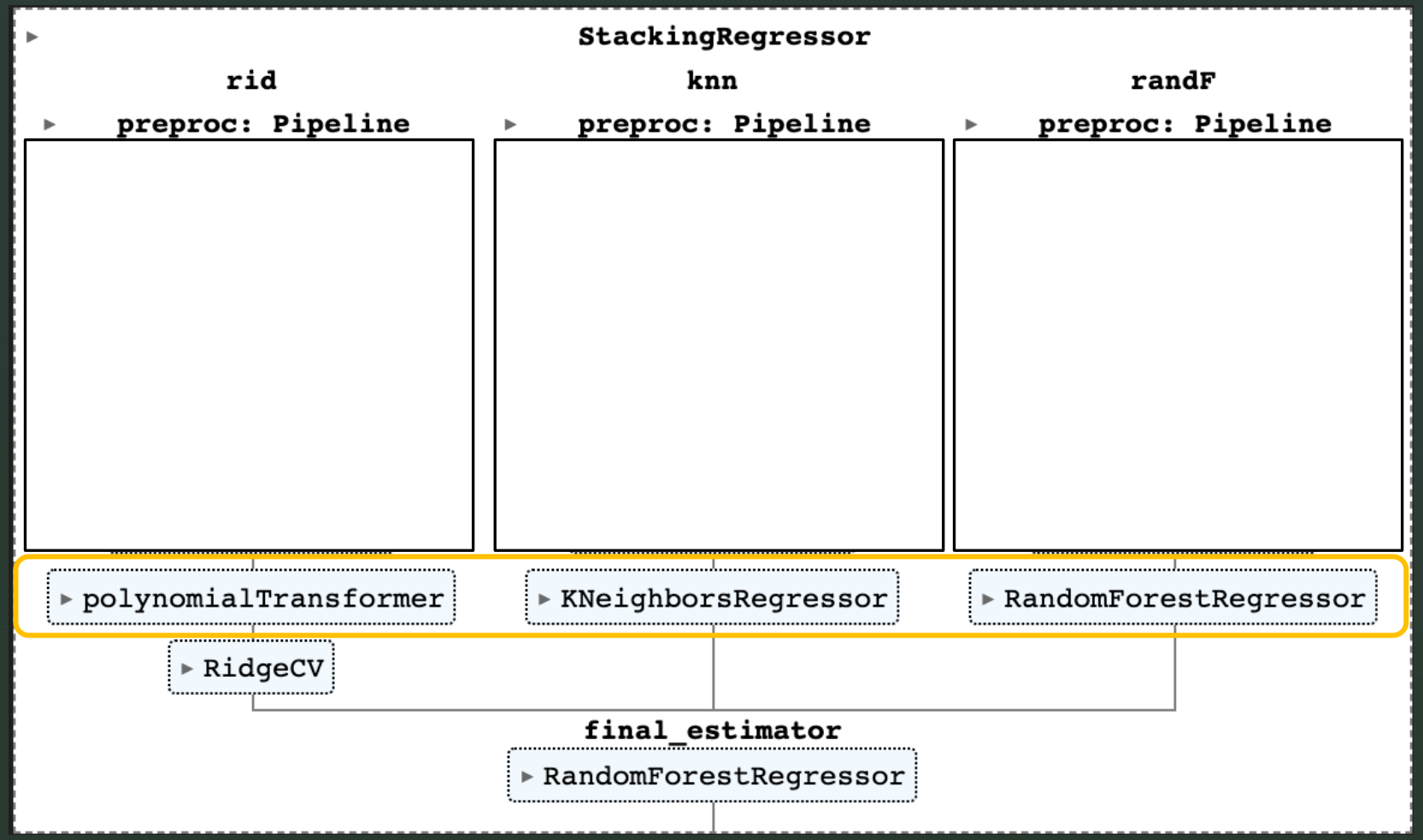
Ensamble Model: Random Forest



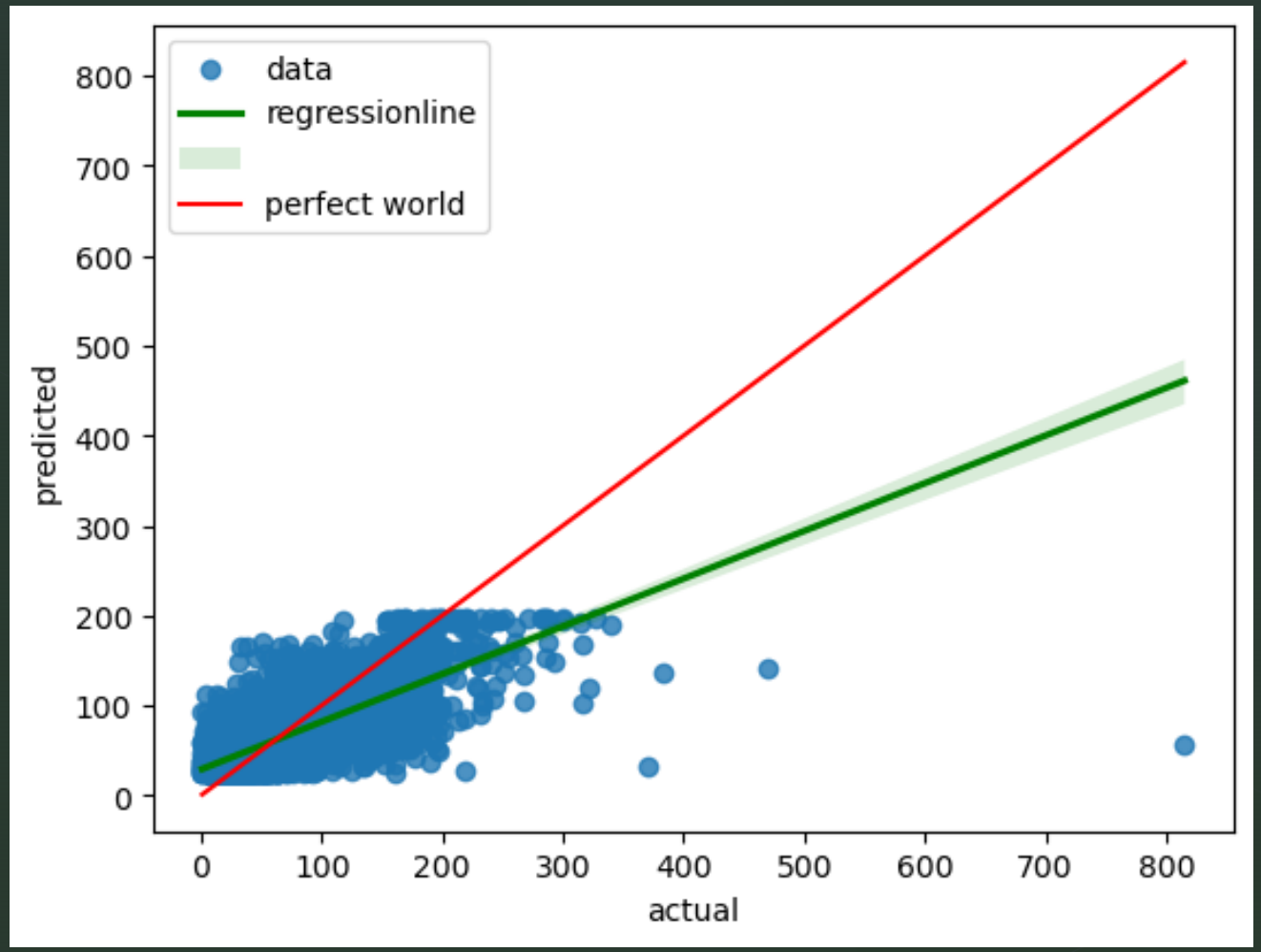
Advanced Ensamble Model: Stacking



Advanced ensamble model: Stacking



Advanced ensamble model: Stacking



Evaluation of our model

Competition Leaderboard

Unless stated otherwise in the Info Page, this leaderboard reflects scores based on only a portion of the total test set until the competition closes. See competition Info for more information.

RANK	USER	PUBLIC SCORE
1	 Sodiq_ University of ibadan	28.59575603
2	 tomy4reel	29.06015845
3	 JefferyJapheth_St	29.11802479

Final Score: RMSE: 31.2
Global Ranking: 46

52		saheedniyi University of lagos	31.94900705
53		chris_somoye Federal university of technology akure	31.99524845
54		Adegite Federal University of Technology Akure	32.07976737

Project Outlook