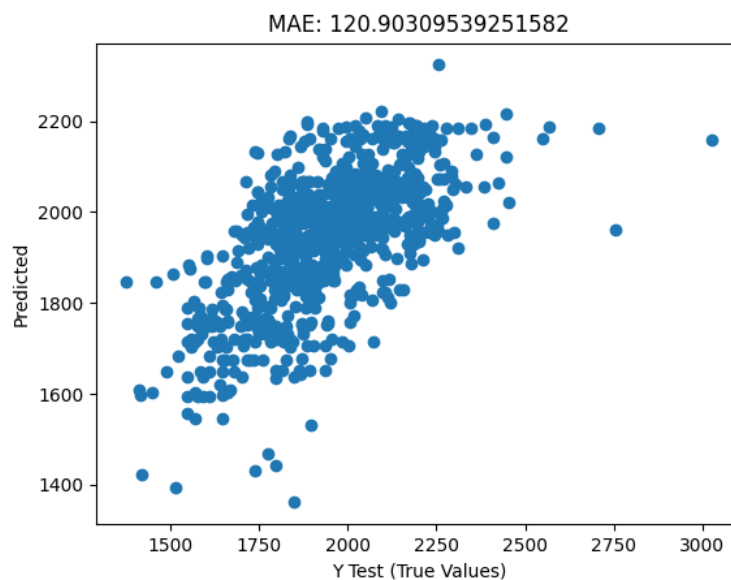


# 1 Predtekmovanje



Slika 1: Graf predikcij v odvisnosti od pravih vrednosti (lokalno testiranje)

## 1.1 Komentar

Atributi, ki sem jih uporabil pri predtekmovanju so naslednji:

- Dnevi tedna - vsak stolpec svoj dan (ponedeljek - nedelja), skupaj 7 atributov
- Časovni intervali v dnevu - vsako uro sem razdelil v 12 intervalov (po 5 min),  $24 \times 12$  skupaj 288 atributov
- Počitnice in dela prosti dnevi - en stolpec za prisotnost počitnic, 1 stolpec za državne praznike, skupaj 2 atributa
- Vreme - povprečna temperatura, količina padavin, prisotnost sneženja, kombinacija dežja in nizke temperature, skupaj 4 atributi

Skupno sem za predtekmovanje uporabil 301 atributov, je pa res, da tisti 4 atributi pri vremenu zelo malo pripomorejo k točnosti napovedovanja.

Zelo pa pripomore treniranje na mrzlih mesecih (januar, februar, november), saj če poznamo testno množico (december) lahko bolje naučimo model.

Uporabljal sem samo model Linear Regression, model Ridge ni pripomogel k boljši predikciji.

## 2 Tekmovanje

Pri napovedovanju vseh linij sem uporabljal iste attribute, le spremenil sem pristop. Namesto grajenja enega modela za vse linije, sem naredil za vsak Route Direction svoj model. Modele sem treniral samo na novembru, če mi je za kakšno linijo zmanjkalo učnih primerov sem za napoved vožnje izbral kar povprečje trajanja ene vožnje skozi celotno leto.

K izboljšanju je nato pripomogla tudi regularizacija modelov. To sem dosegel z modelom Ridge.

Z omenjenim pristopom sem na tekmovanju dosegel 11. mesto (MAE 162.80021).