

# WB XAI-2 PD5

Jakub Szypuła

09/05/2021

## PDP dla wybranych zmiennych

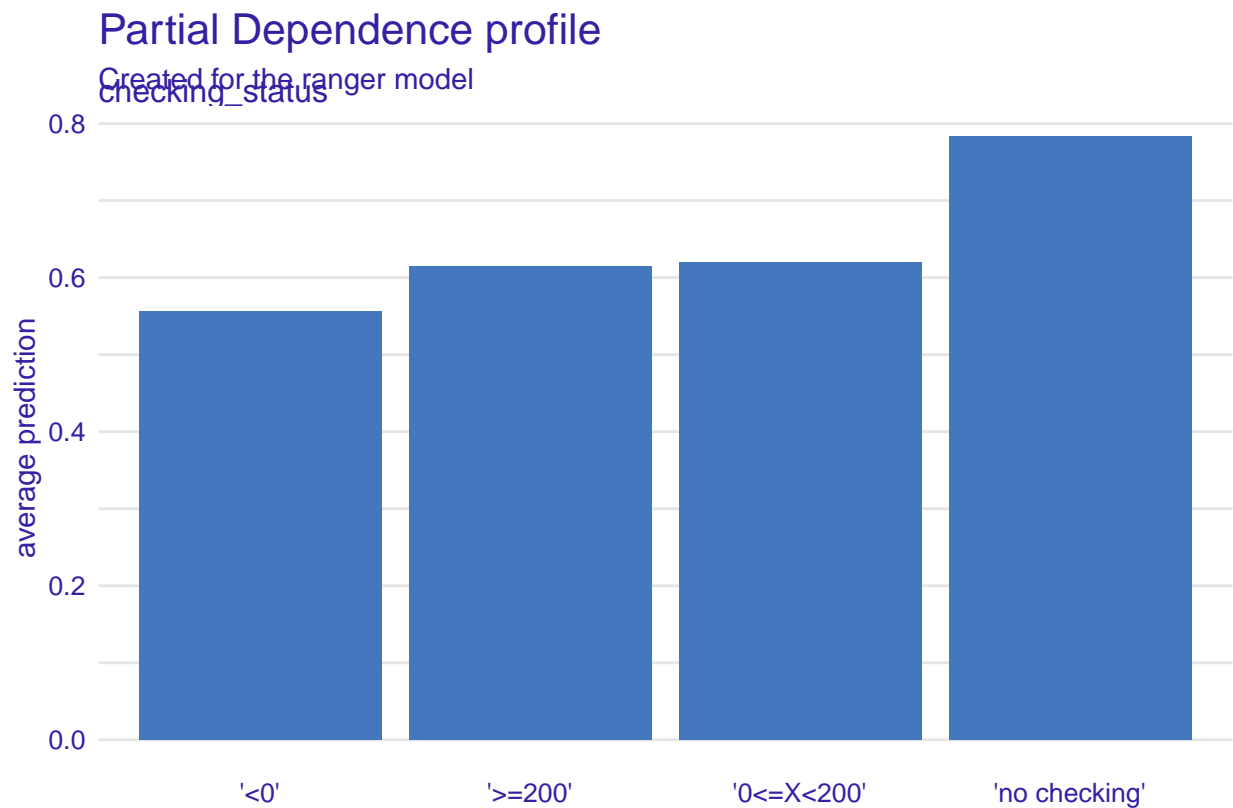
Na podstawie doświadczeń z poprzednich PD postanowiłem wybrać zachowanie 4 zmiennych (kryterium była ważność zmiennych):

- checking\_status
- duration
- credit\_amount
- age

```
pdp_1 <- model_profile(explainer, variables = "checking_status")
```

```
## 'variable_type' changed to 'categorical' due to lack of numerical variables.
```

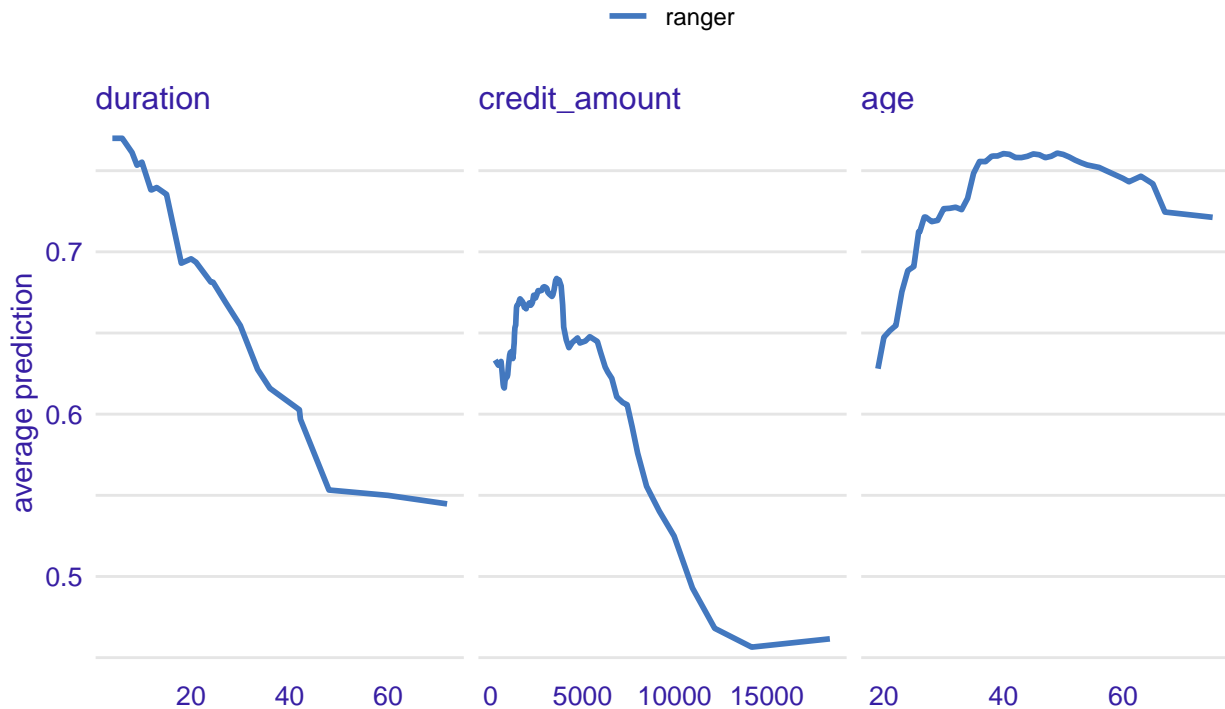
```
plot(pdp_1)
```



```
pdp_2 <- model_profile(explainer, variables = "duration")
pdp_3 <- model_profile(explainer, variables = "credit_amount")
pdp_4 <- model_profile(explainer, variables = "age")
plot(pdp_2, pdp_3, pdp_4)
```

## Partial Dependence profile

Created for the ranger model



Powtarza się po raz kolejny to co się działo na przestrzeni poprzednich prac domowych, a mianowicie widać zależność większego prawdopodobieństwa pozytywnej oceny kredytobiorcy przy większym rachunku bieżącym, spadku wraz ze wzrostem długości i wysokości kredytu oraz niższego prawdopodobieństwa dla niskich i wysokich wartości zmiennej age.

Po tym wszystkim można wyciągnąć wniosek, że taka zależność na pewno istnieje w modelu i być może jest dobrą heurystyką przy wyznaczaniu prawdopodobieństwa spłaty kredytu.

## ALE dla wybranych zmiennych

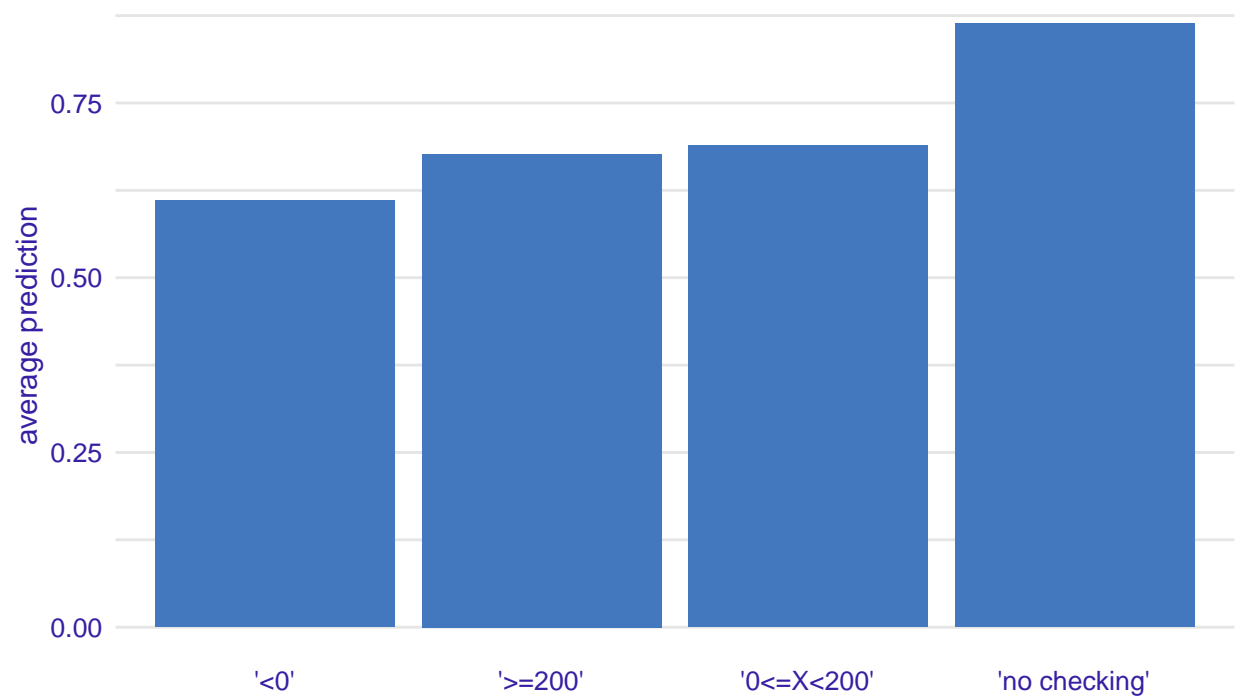
```
ale_1 <- model_profile(explainer, variables = "checking_status", type = "accumulated")
```

```
## 'variable_type' changed to 'categorical' due to lack of numerical variables.
```

```
plot(ale_1)
```

## Accumulated Dependence profile

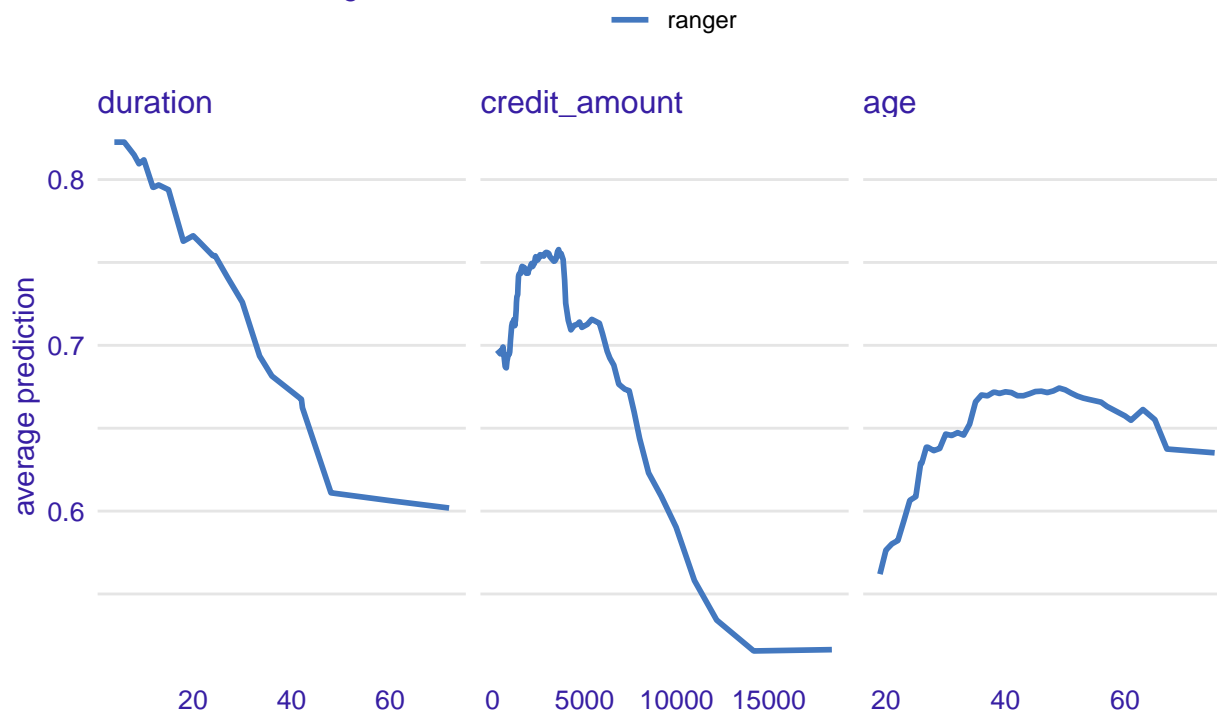
Created for the ranger model  
checking\_status



```
ale_2 <- model_profile(explainer, variables = "duration", type = "accumulated")
ale_3 <- model_profile(explainer, variables = "credit_amount", type = "accumulated")
ale_4 <- model_profile(explainer, variables = "age", type = "accumulated")
plot(ale_2, ale_3, ale_4)
```

## Accumulated Dependence profile

Created for the ranger model



Wykresy wyglądają prawie identycznie jak w przypadku PDP, z drobnym wyjątkiem - ta metoda pokazuje większe średnie przewidywane prawdopodobieństwo dla dużych wartości zmiennej `duration` i mniejsze dla `credit_amount` w porównaniu do metody PDP. Można się zastanowić, czy to nie kwestia tego, że ponieważ predykcja maleje dla wzrostu obu zmiennych, a istnieje pewnie pewna korelacja dla dużych wartości (duże kredyty będą spłacane dłużej), więc niższa wartość wynikająca z jednej zmiennej będzie rekompensowana przez drugą. Można się zastanowić, jak postarać się naprawić, być może rata pożyczki byłaby lepsza, aczkolwiek traci się wtedy informację na temat tego jak długi będzie faktycznie ten kredyt.

## Różne kombinacje rozmiaru siatki

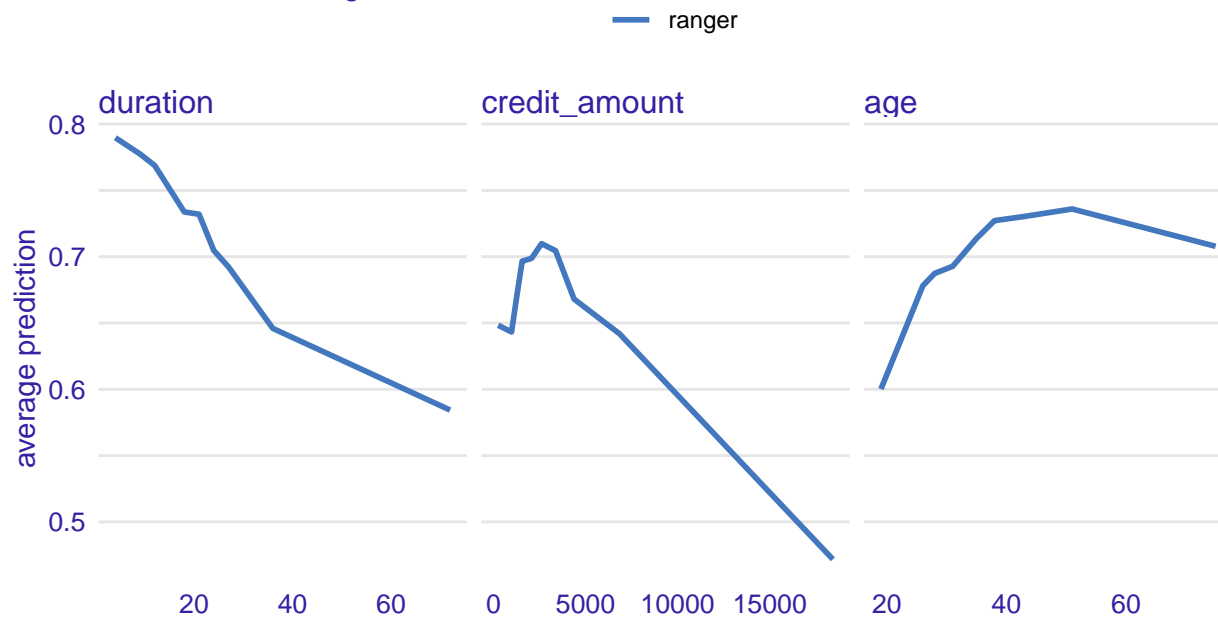
Ponieważ ten punkt dotyczy krzywych, pominę zmienną kategorię (czyli `Checking_status`).

### PDP

```
pdp_2 <- model_profile(explainer, variables = "duration", N = 10, grid_points = 10)
pdp_3 <- model_profile(explainer, variables = "credit_amount", N = 10, grid_points = 10)
pdp_4 <- model_profile(explainer, variables = "age", N = 10, grid_points = 10)
plot(pdp_2, pdp_3, pdp_4)
```

## Partial Dependence profile

Created for the ranger model



```
pdp_2 <- model_profile(explainer, variables = "duration", N = 1000, grid_points = 10)
pdp_3 <- model_profile(explainer, variables = "credit_amount", N = 1000, grid_points = 10)
pdp_4 <- model_profile(explainer, variables = "age", N = 1000, grid_points = 10)
plot(pdp_2, pdp_3, pdp_4)
```

## Partial Dependence profile

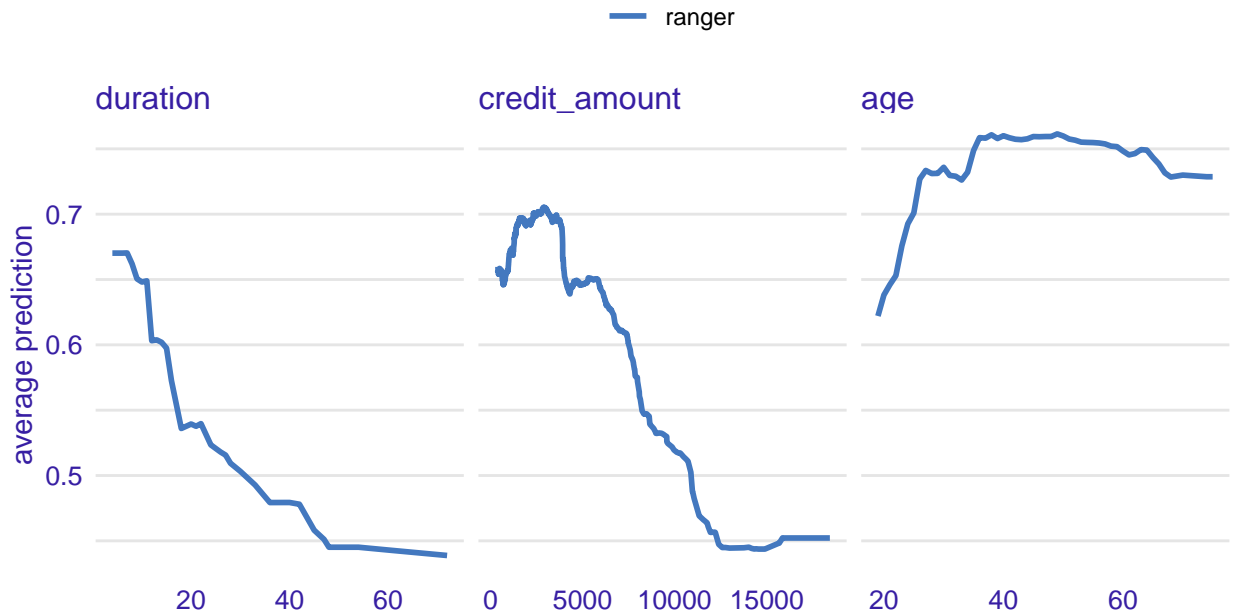
Created for the ranger model



```
pdp_2 <- model_profile(explainer, variables = "duration", N = 10, grid_points = 1000)
pdp_3 <- model_profile(explainer, variables = "credit_amount", N = 10, grid_points = 1000)
pdp_4 <- model_profile(explainer, variables = "age", N = 10, grid_points = 1000)
plot(pdp_2, pdp_3, pdp_4)
```

## Partial Dependence profile

Created for the ranger model



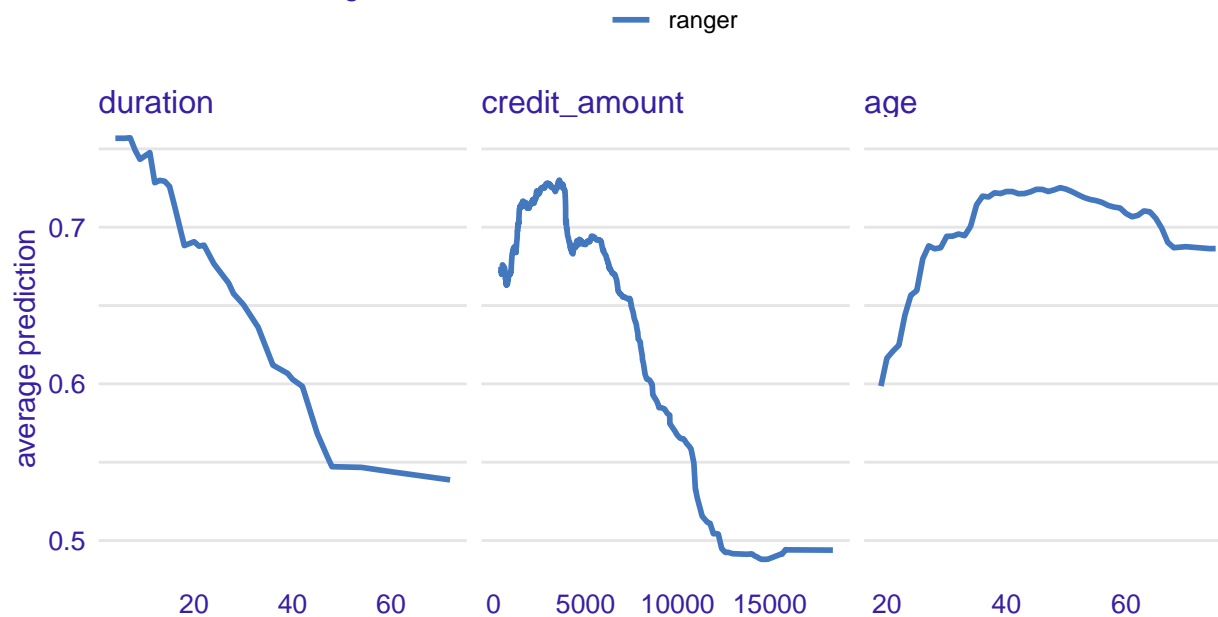
```
pdp_2 <- model_profile(explainer, variables = "duration", N = 1000, grid_points = 1000)
pdp_3 <- model_profile(explainer, variables = "credit_amount", N = 1000, grid_points = 1000)

## Aggregating predictions.. Progress: 90%. Estimated remaining time: 3 seconds.

pdp_4 <- model_profile(explainer, variables = "age", N = 1000, grid_points = 1000)
plot(pdp_2, pdp_3, pdp_4)
```

## Partial Dependence profile

Created for the ranger model

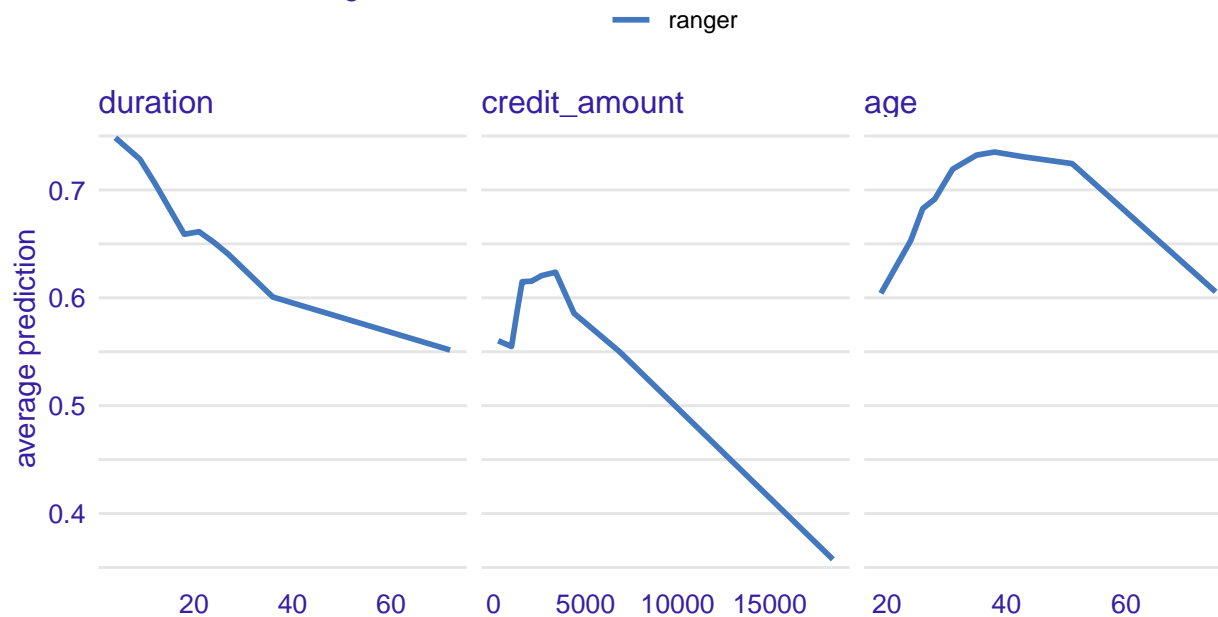


### ALE

```
ale_2 <- model_profile(explainer, variables = "duration", type = "accumulated", N = 10, grid_points = 10)
ale_3 <- model_profile(explainer, variables = "credit_amount", type = "accumulated", N = 10, grid_points = 10)
ale_4 <- model_profile(explainer, variables = "age", type = "accumulated", N = 10, grid_points = 10)
plot(ale_2, ale_3, ale_4)
```

## Accumulated Dependence profile

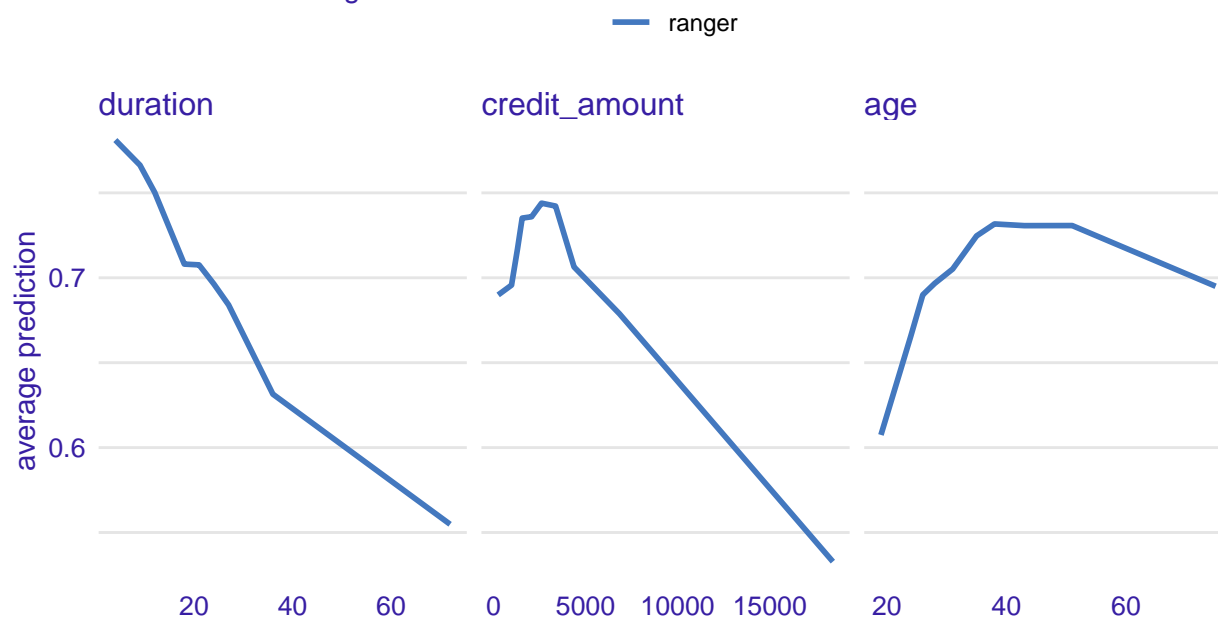
Created for the ranger model



```
ale_2 <- model_profile(explainer, variables = "duration", type = "accumulated", N = 1000, grid_points = 100)
ale_3 <- model_profile(explainer, variables = "credit_amount", type = "accumulated", N = 1000, grid_points = 100)
ale_4 <- model_profile(explainer, variables = "age", type = "accumulated", N = 1000, grid_points = 100)
plot(ale_2, ale_3, ale_4)
```

## Accumulated Dependence profile

Created for the ranger model

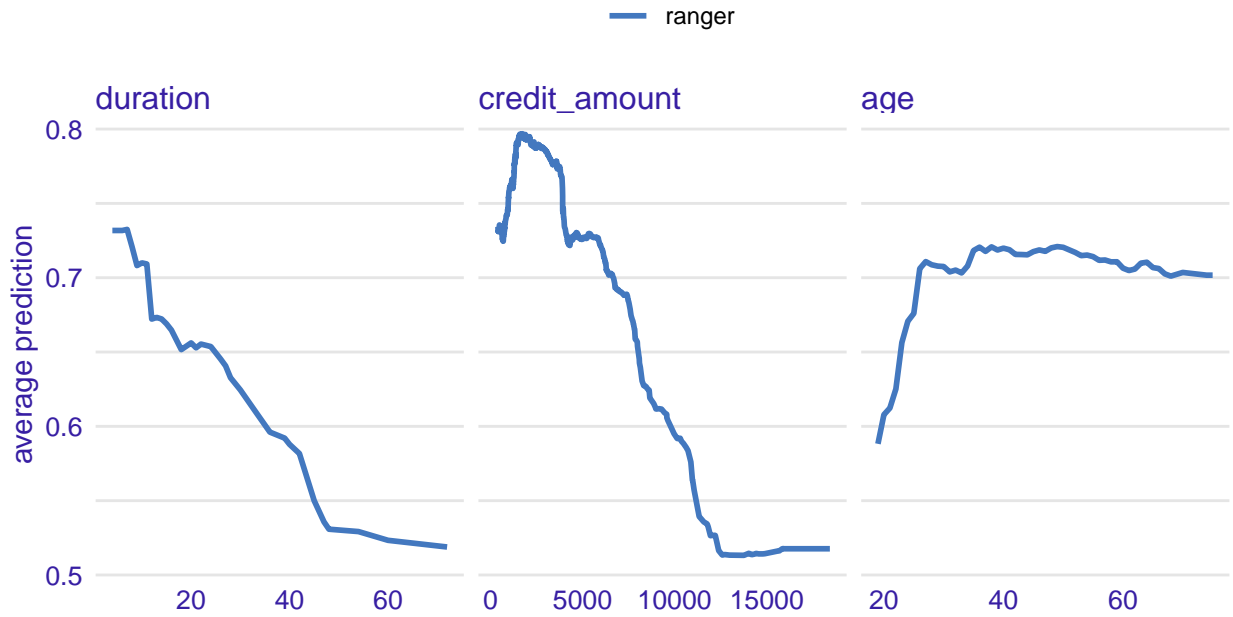




```
ale_2 <- model_profile(explainer, variables = "duration", type = "accumulated", N = 10, grid_points = 100)
ale_3 <- model_profile(explainer, variables = "credit_amount", type = "accumulated", N = 10, grid_points = 100)
ale_4 <- model_profile(explainer, variables = "age", type = "accumulated", N = 10, grid_points = 1000)
plot(ale_2, ale_3, ale_4)
```

## Accumulated Dependence profile

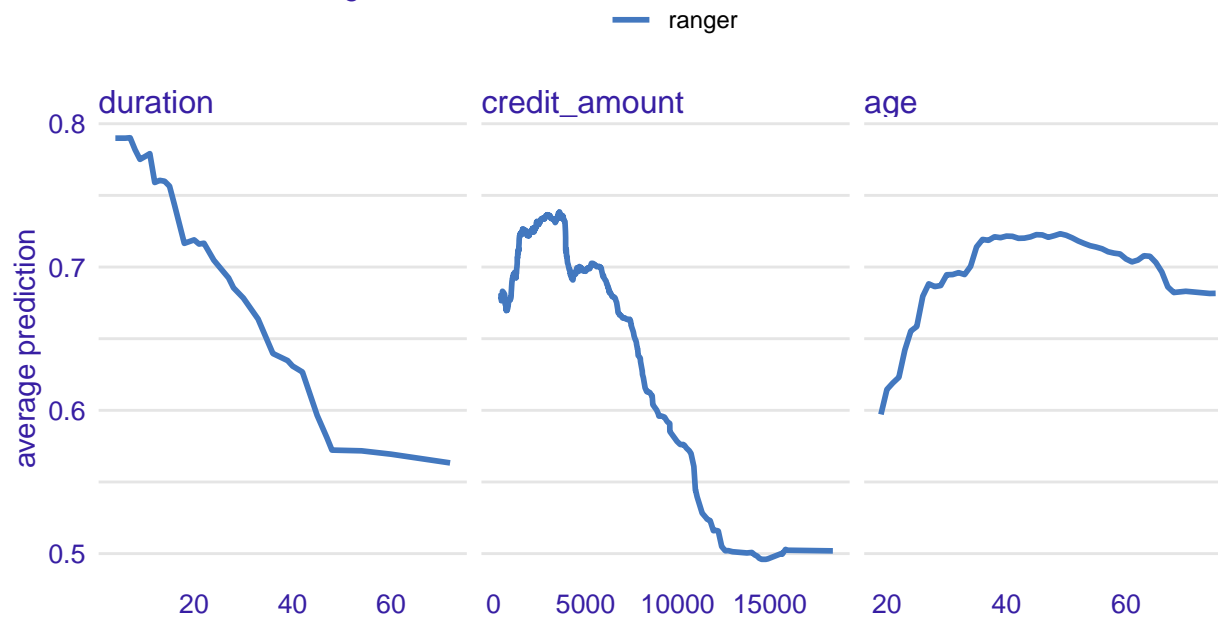
Created for the ranger model



```
ale_2 <- model_profile(explainer, variables = "duration", type = "accumulated", N = 1000, grid_points = 100)
ale_3 <- model_profile(explainer, variables = "credit_amount", type = "accumulated", N = 1000, grid_points = 100)
ale_4 <- model_profile(explainer, variables = "age", type = "accumulated", N = 1000, grid_points = 1000)
plot(ale_2, ale_3, ale_4)
```

## Accumulated Dependence profile

Created for the ranger model



Jak widać zmiana siatki potrafi powodować duże zmiany, zwłaszcza na małe wartości. Zwiększanie wartości nie zmienia jednak znacząco kształtu i wartości wykresów, co pozwala wyciągnąć wniosek, że metoda jest bardzo wrażliwa przy stosowaniu ogólnej siatki (zmieniają się maksymalne i minimalne wartości). Rozkład punktów też ma mniejsze znaczenie niż rozmiar siatki.