



**Wydział Matematyki  
i Nauk Informatycznych**

POLITECHNIKA WARSZAWSKA



# CONSUMER LOANS

## DEFAULT

## PREDICTOR



**JULIA KRUK, MICHAŁ PIECHOTA**



# CEL BIZNESOWY

---

- PRZEWIDZIEĆ, CZY POŻYCZKOBIORCA BĘDZIE MIAŁ ZALEGŁOŚCI W SPŁACIE POŻYCZKI
- PRZEWIDUJEMY CECHE **DEFAULT FLAG**



# DANE

- O ŻYCIU KLIENTA, NP. AGE, MARITAL STATUS, EDUCATION
- O PRACY KLIENTA, NP. INCOME, WORK SENIORITY, ECONOMIC SECTOR
- O FUNDUSZACH KLIENTA, NP. DEBIT CARD, SAVING ACCOUNT, DEPOSIT



# KORELACJE

## ➤ MIĘDZY CECHAMI

NO_OF_DEPENDENTS	HOUSEHOLD_MEMBERS
CURRENT_ACCOUNT	DEBIT_CARD
Owner without mortgage	Living with family
married	HOUSEHOLD_MEMBERS

## ➤ Z TARGETEM

- MARITAL\_STATUS\_single
- MARITAL\_STATUS\_married
- RESIDENTIAL\_PLACE\_Living with family
- HOUSEHOLD\_MEMBERS



# FEATURE IMPORTANCE

## ➤ NAJWAŻNIEJSZE CECHY

- INCOME
- AGE
- BUSINESS AGE
- WORK SENIORITY
- EMPLOYEE NO

## ➤ NAJMNIEJ WAŻNE

- ECONOMIC\_SECTOR
  - Professional, scientific and technical activities
  - Education
  - Information and communication
- RESIDENTIAL\_PLACE\_Owner with mortgage

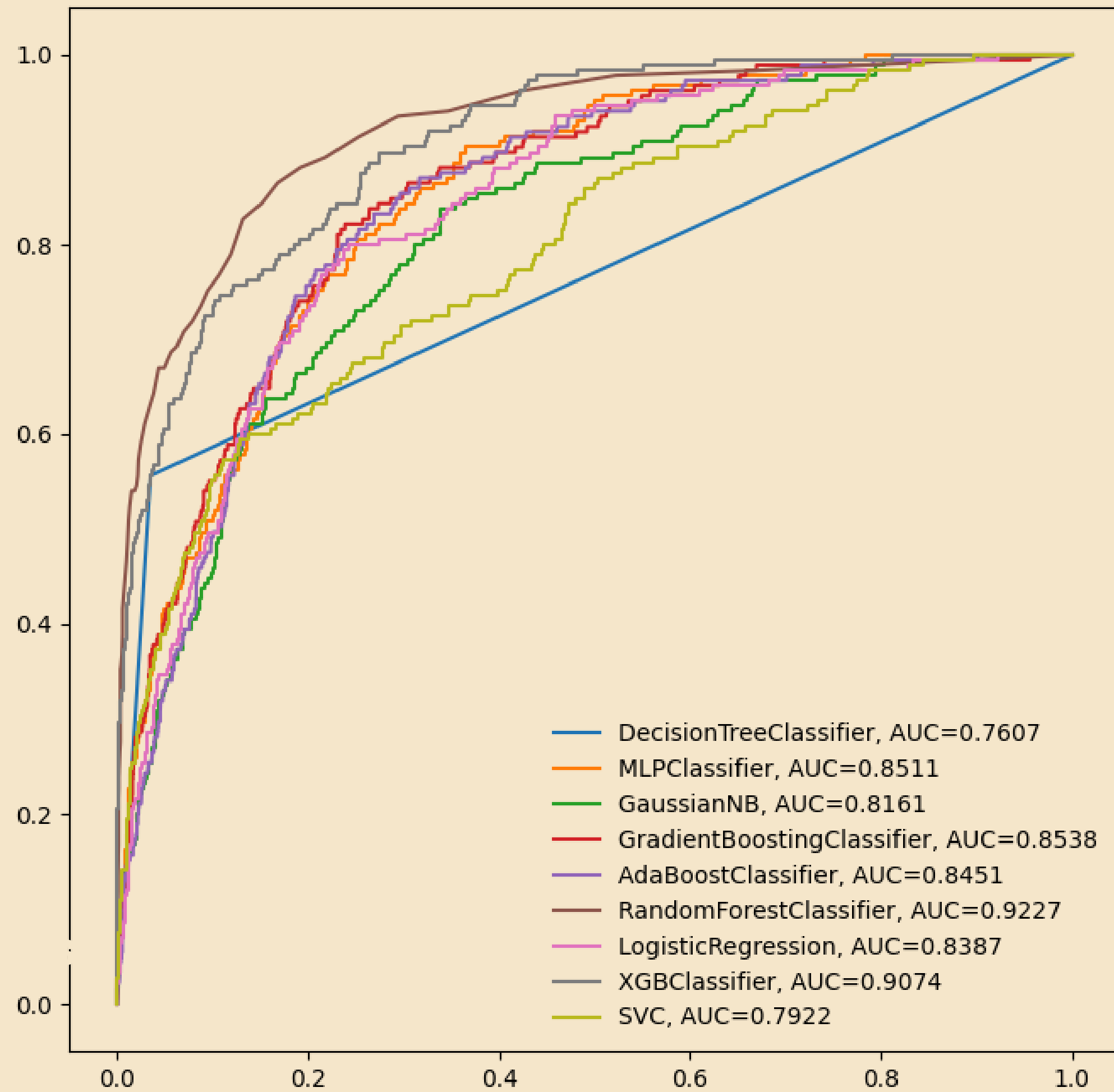


# WYBÓR MODELI

**ZBIÓR DANYCH BYŁ NIEZBALANSOWANY,  
ZATEM:  
NAJWAŻNIEJSZA MIARA - AUC (GINI)**



ROC Curve for Different Models (Validation Data)





# NAJLEPSZE MODELE - GINI

RANDOM FOREST CLASSIFIER

**0.840**

XGBOOST CLASSIFIER

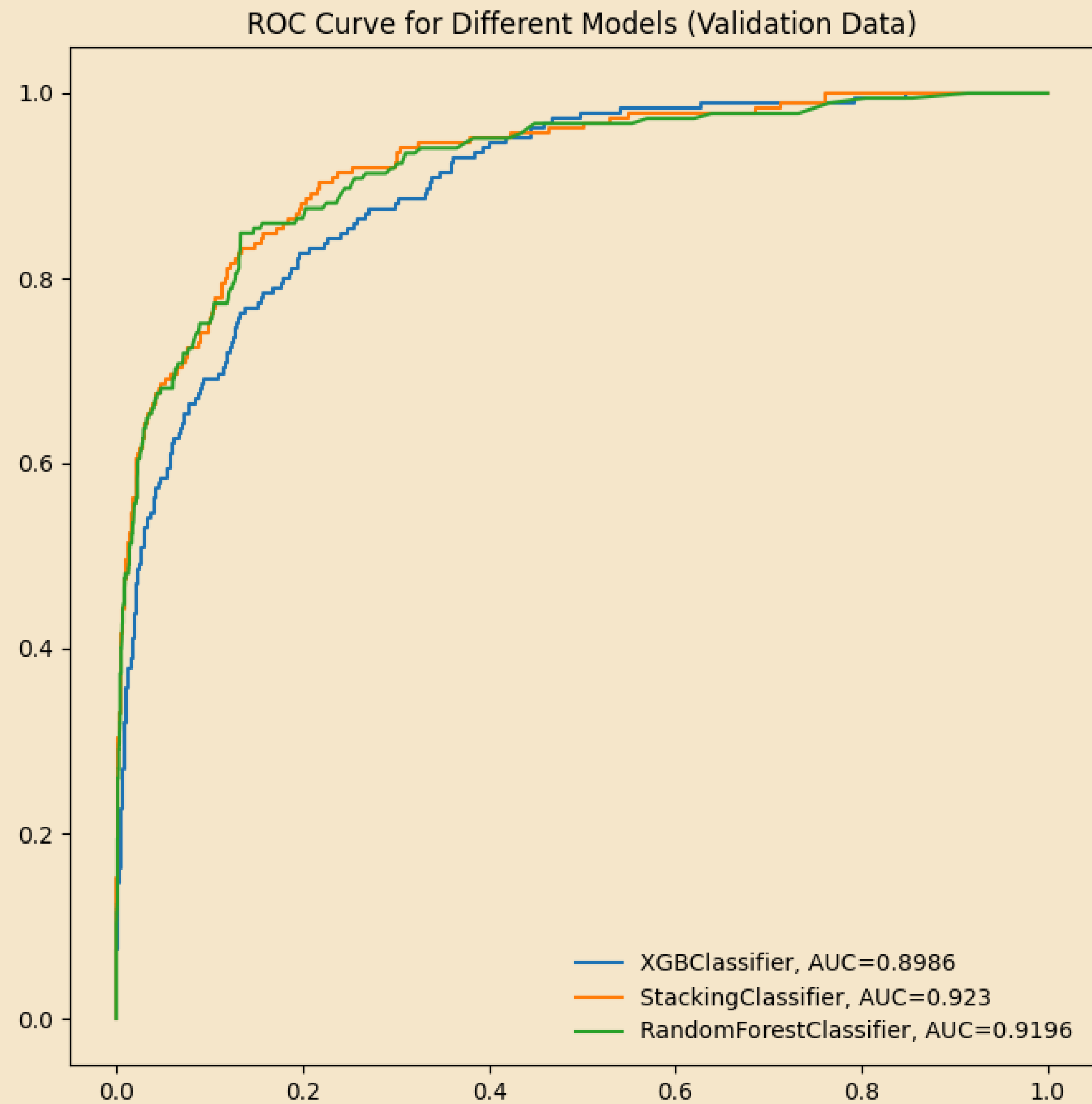
**0.797**

STACKING CLASSIFIER  
(RF + XGB)

**0.846**

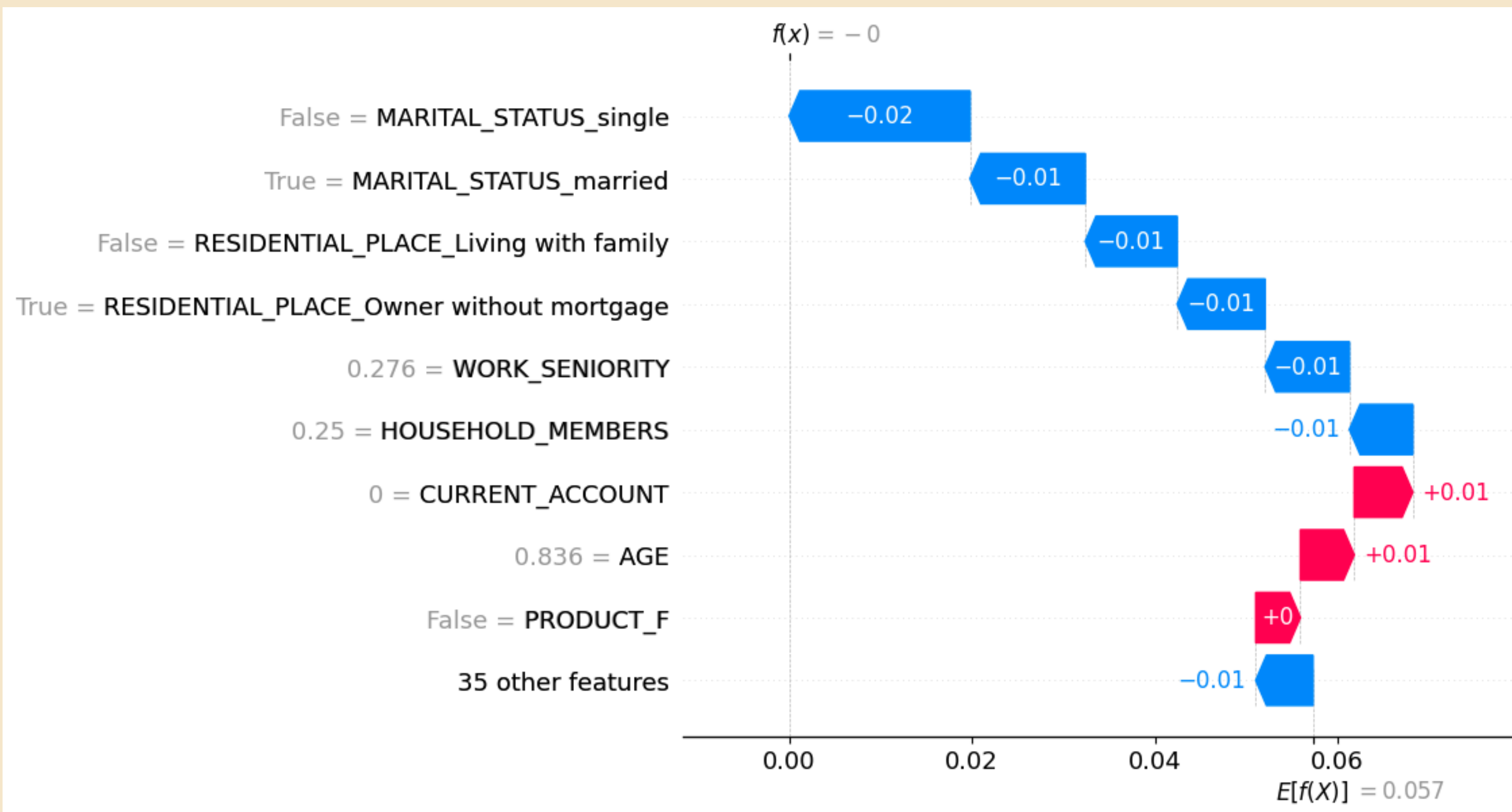
**WYBIERAMY RANDOM FOREST**





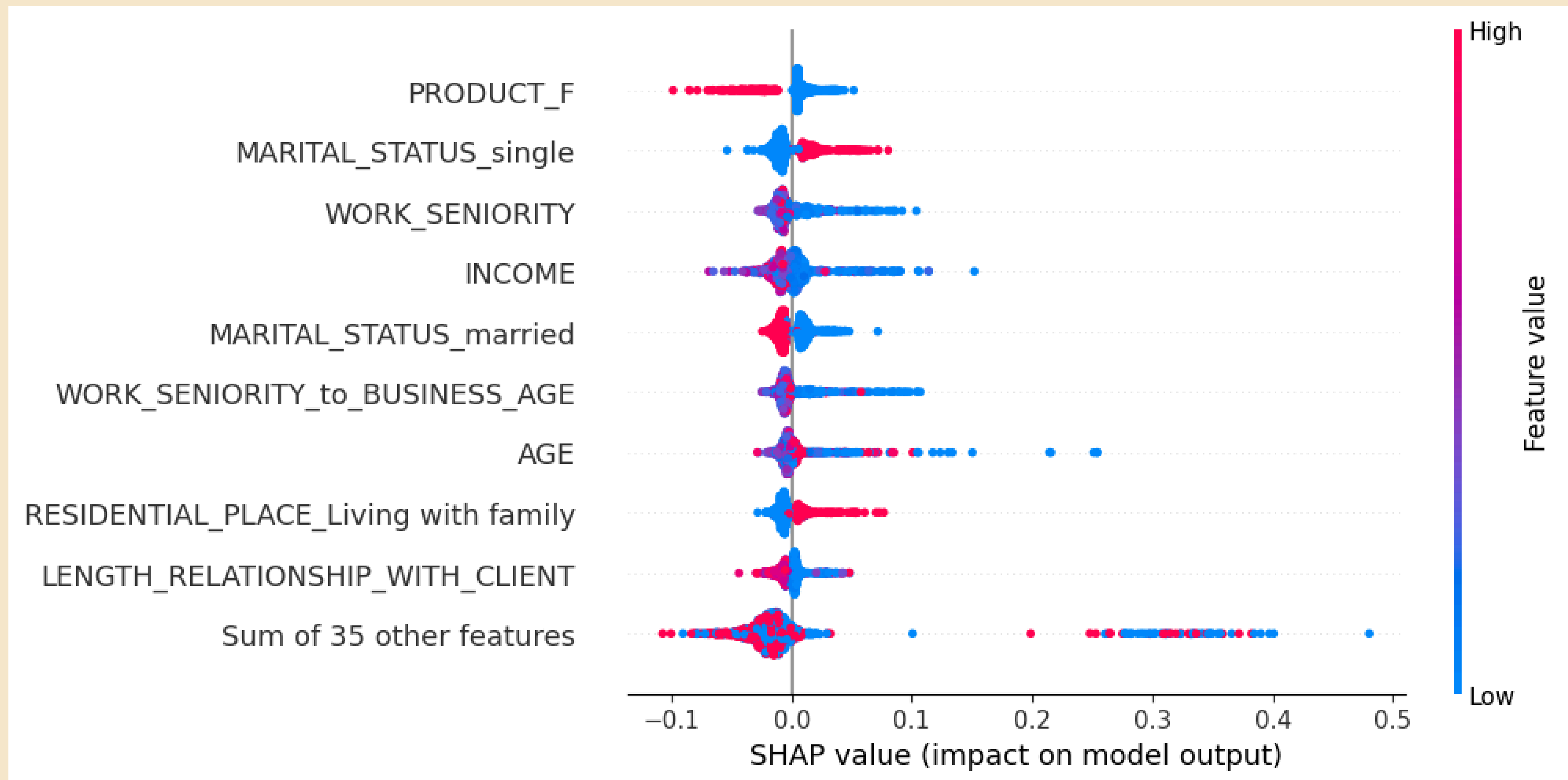


# DZIAŁANIE MODELU RANDOM FOREST



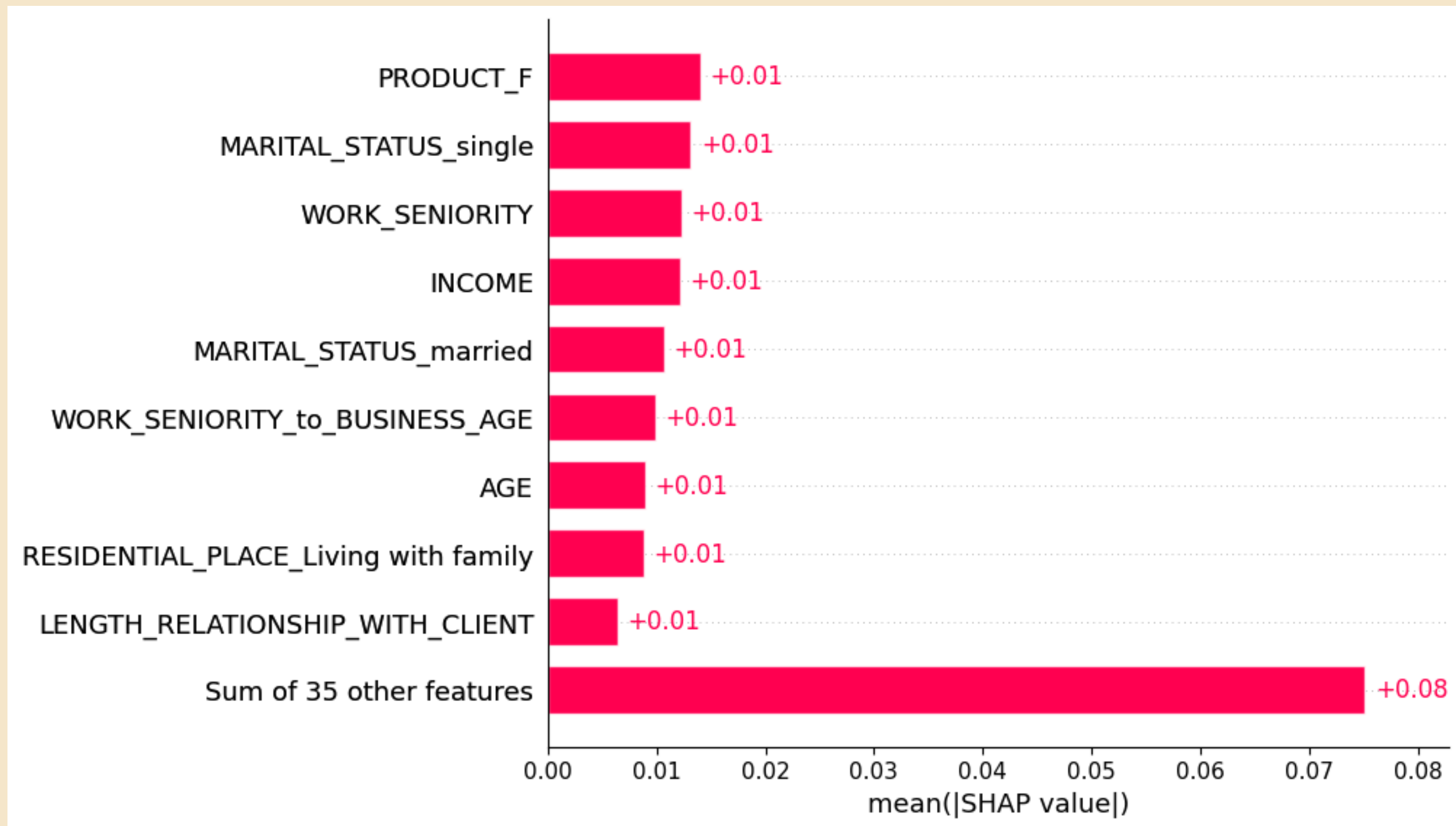


# DZIAŁANIE MODELU RANDOM FOREST





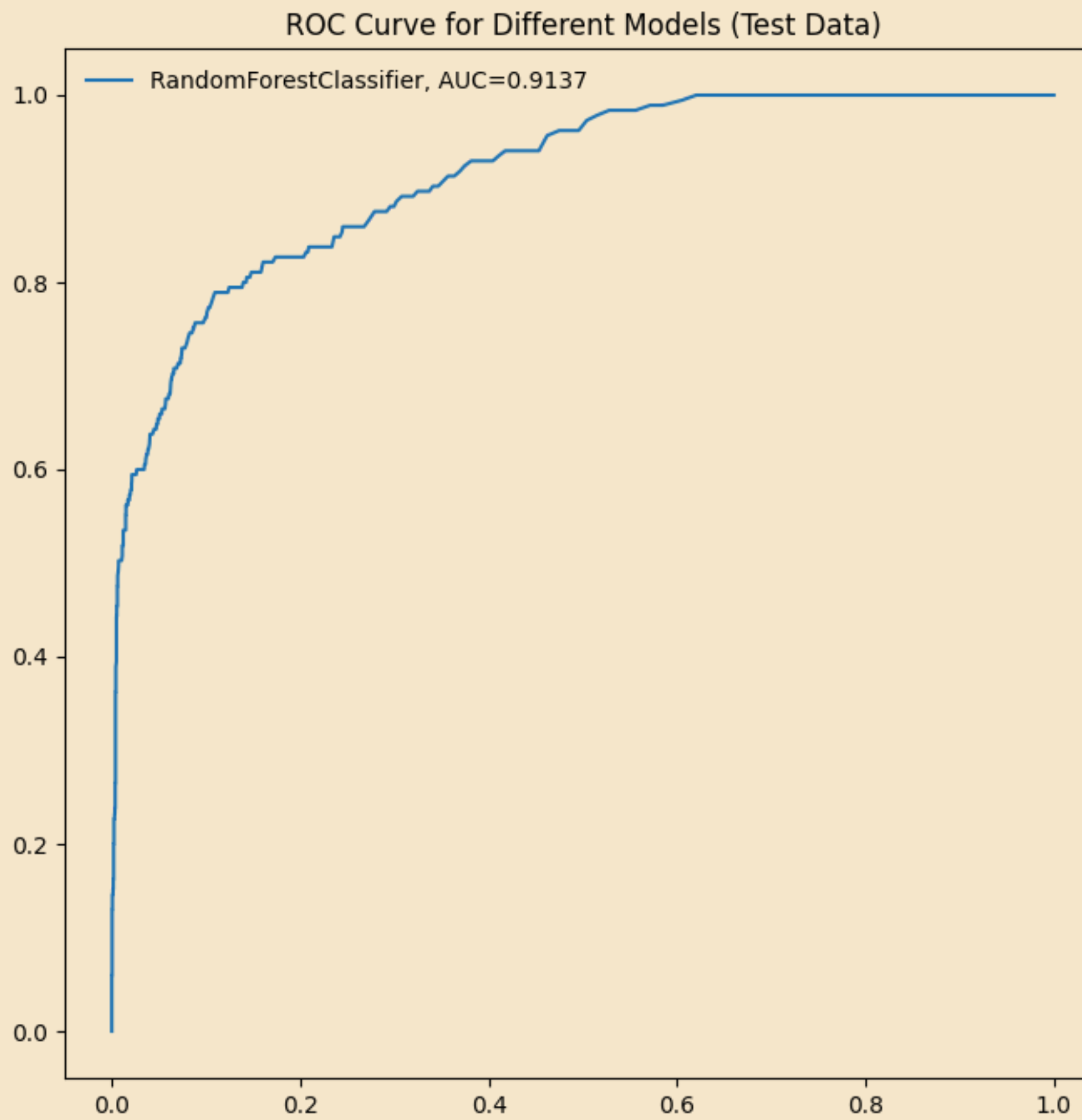
# DZIAŁANIE MODELU RANDOM FOREST





# WYNIKI WALIDACJI

**GINI: 0.823**





**Wydział Matematyki  
i Nauk Informatycznych**

POLITECHNIKA WARSZAWSKA



# CONSUMER LOANS

## DEFAULT

## PREDICTION



**JULIA KRUK, MICHAŁ PIECHOTA**