

# Analiza tunowalności algorytmów uczenia maszynowego

Bartosz Maj, Bartosz Olszewski, Jan Kruszewski



# Wybrane algorytmy uczenia maszynowego

- Las losowy - biblioteka sklearn
- Regresja logistyczna - biblioteka sklearn
- XGBoost - biblioteka xgboost



# Wybrane zbiory danych

- Wine - zawiera cechy charakterystycznych dla danego wina wraz z jego ogólną oceną
- Drug - posiada dane medyczne dla pacjentów oraz lek, który został podany
- Iris - zawiera dane opisujące wymiary irysów oraz nazwę konkretnego gatunku
- Titanic - posiada dane opisujące osoby znajdujące się na statku Tytanic razem z informacją, czy dana osoba przeżyła



# Wybrane hiperparametry



# Wybrane hiperparametry

## Las losowy

### Wartości hiper-parametrów dla metod Grid Search, Random Search oraz Bayes Search

- `n_estimators` - określa liczbę drzew decyzyjnych tworzonych w modelu.
- `max_depth` - ustala maksymalną głębokość każdego drzewa w modelu.
- `min_samples_split` - definiuje minimalną liczbę próbek wymaganą do podziału węzła wewnętrznego.
- `min_samples_leaf` - określa minimalną liczbę próbek, które muszą znaleźć się w liście drzewa.



# Wybrane hiperparametry

## Regresja logistyczna

### Wartości hiper-parametrów dla metod Grid Search, Random Search oraz Bayes Search

- C - parametr regularyzacji kontrolujący siłę karania za złożoność modelu; niższe wartości oznaczają silniejszą regularyzację.
- penalty - określa rodzaj funkcji karnej używanej w modelu
- solver - wybiera algorytm optymalizacji używany do dopasowania modelu



# Wybrane hiperparametry

## XGBoost

### Wartości hiper-parametrów dla metod Grid Search, Random Search oraz Bayes Search

- `n_estimators` - określa liczbę iteracji boostingowych, czyli liczbę drzew tworzonych w modelu.
- `max_depth` - ustala maksymalną głębokość każdego drzewa.
- `learning_rate` - określa wielkość kroku w każdej iteracji boostingowej. Niższe wartości spowalniają proces uczenia, co może poprawić dokładność.
- `subsample` - określa odsetek próbek losowo wybieranych do treningu każdego drzewa, co pomaga zapobiegać przeuczeniu modelu.

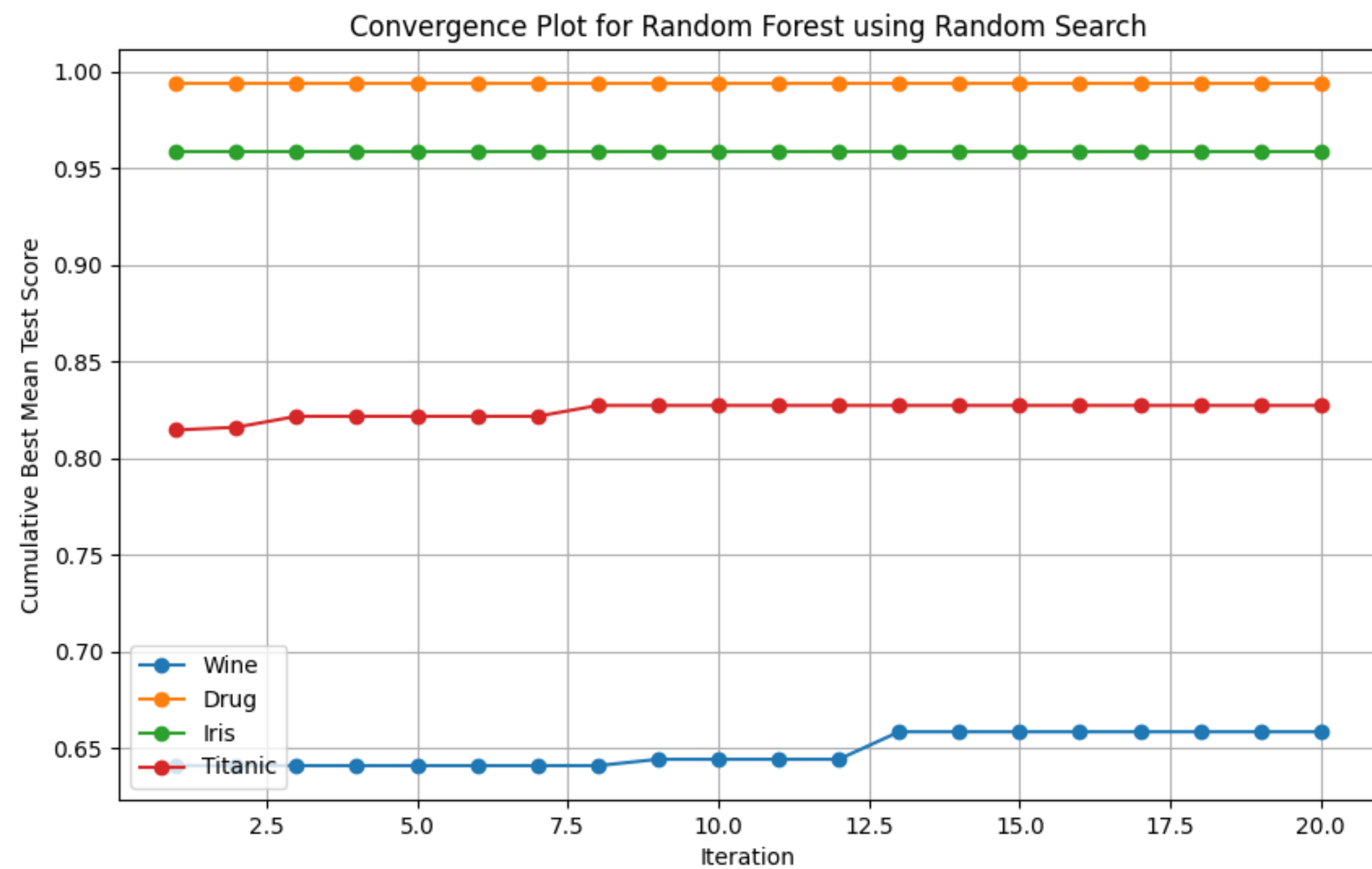


**Wymagana liczba iteracji**



# Wymanagna liczba iteracji

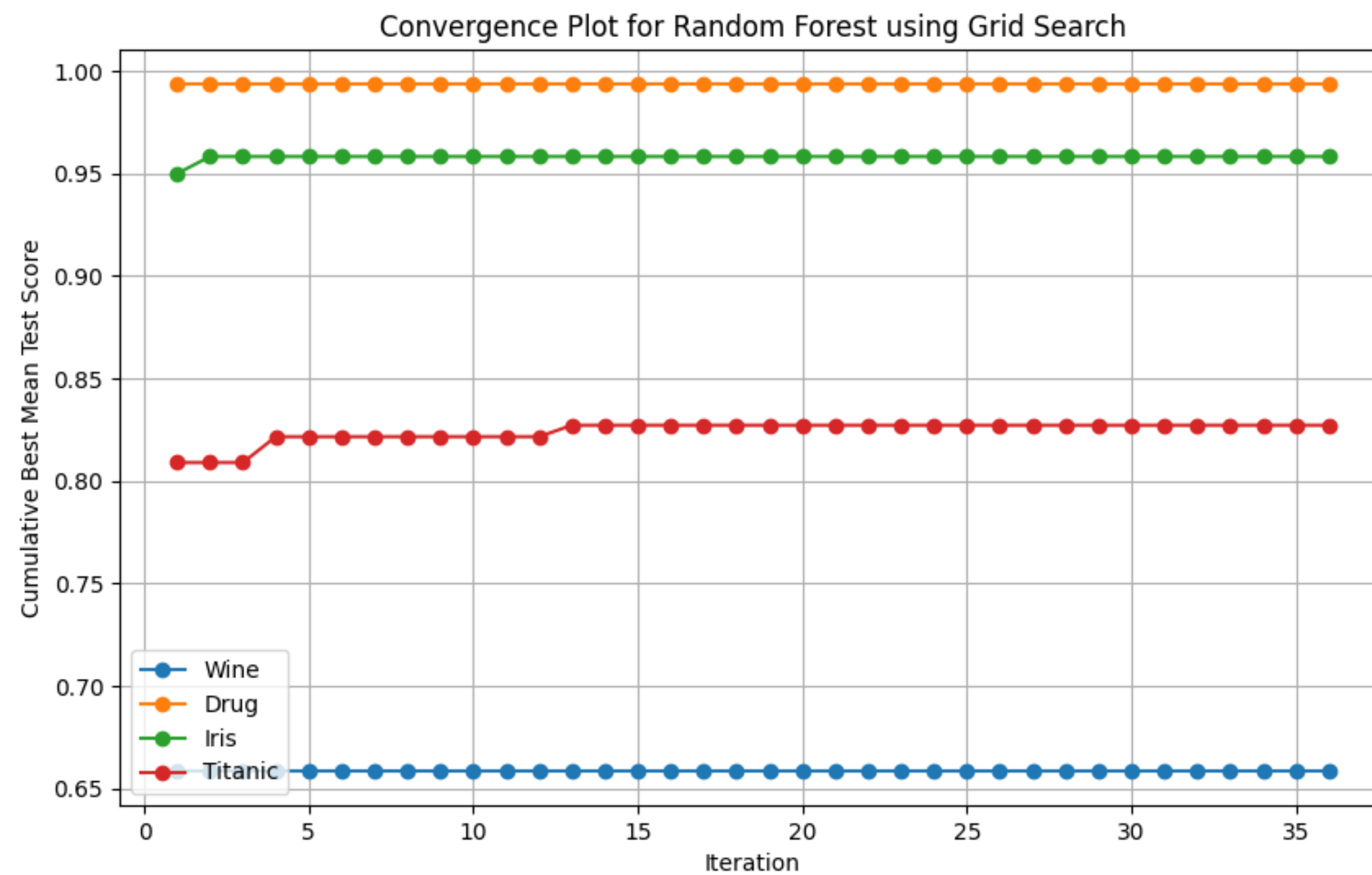
## Las losowy





# Wymanagna liczba iteracji

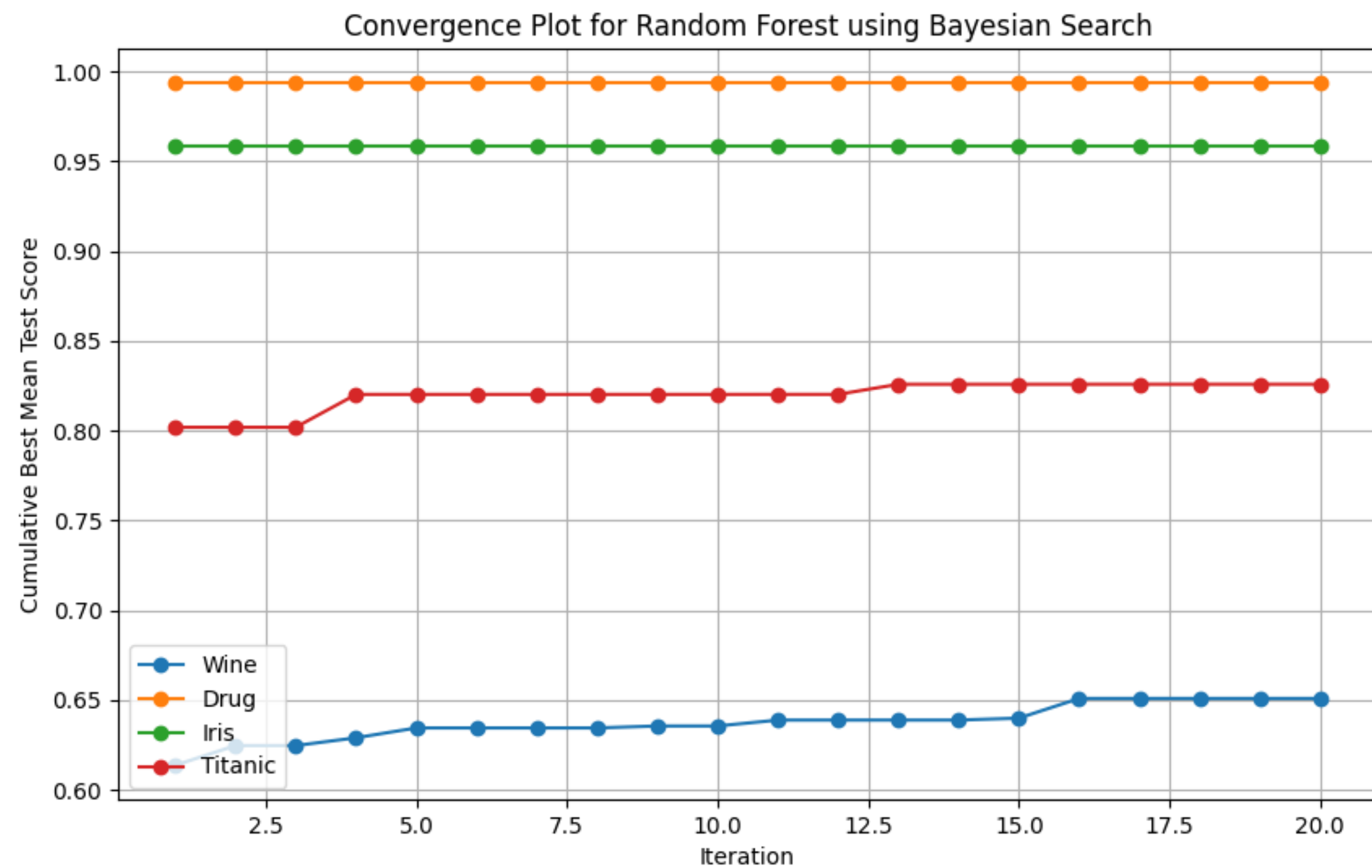
## Las losowy





# Wymanagna liczba iteracji

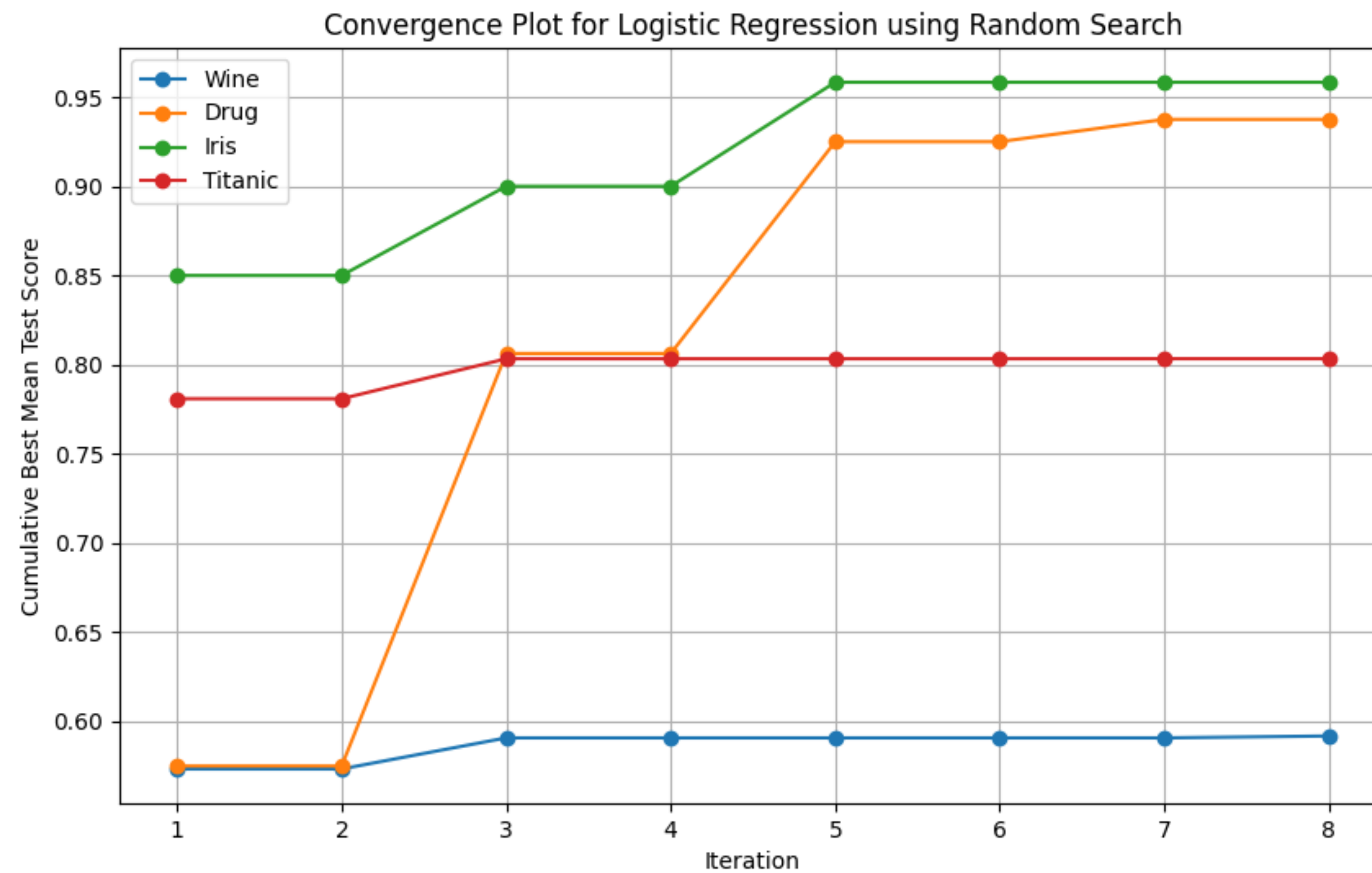
## Las losowy





# Wymanagna liczba iteracji

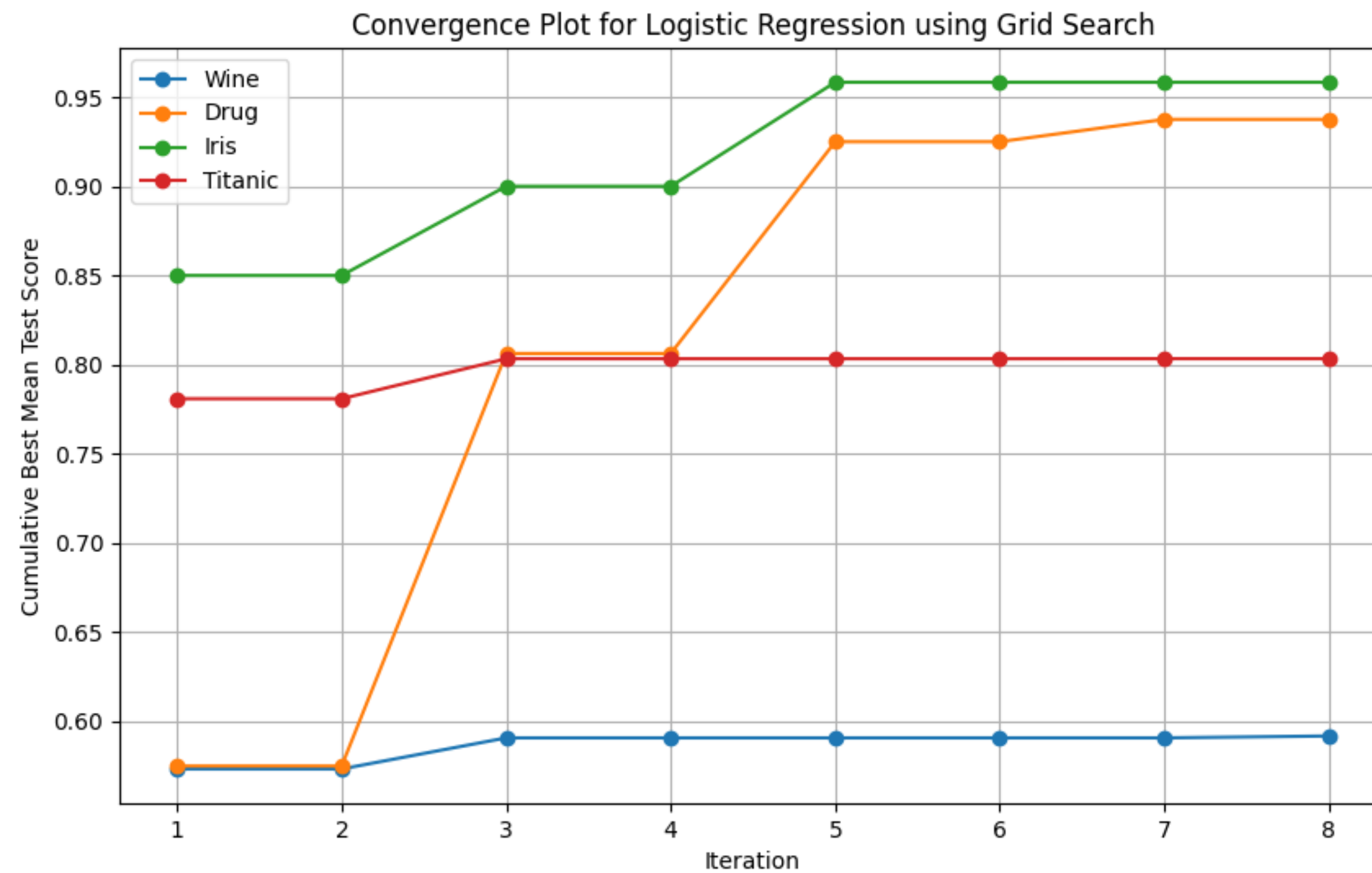
## Regresja logistyczna





# Wymanagna liczba iteracji

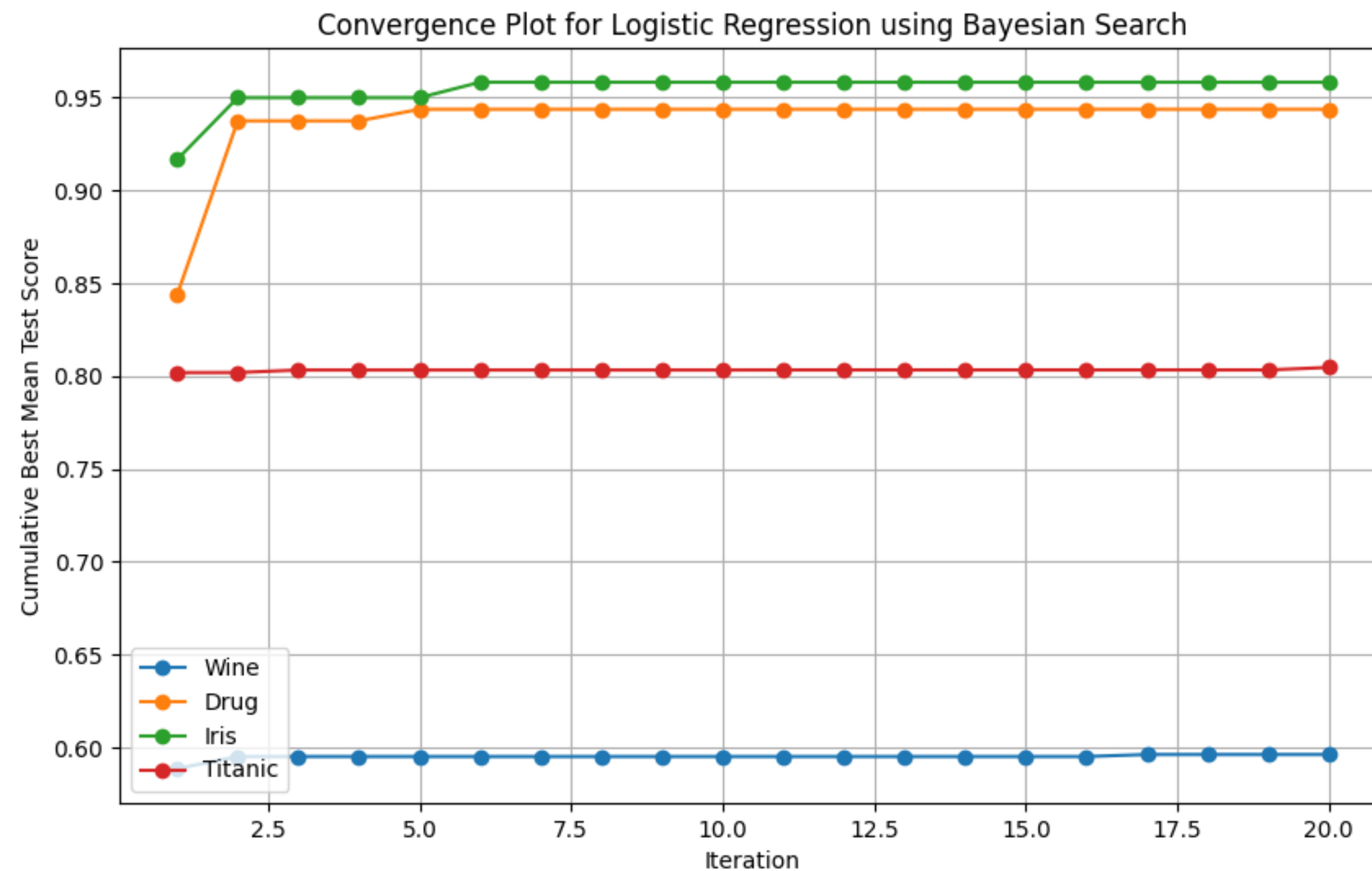
## Regresja logistyczna





# Wymanagna liczba iteracji

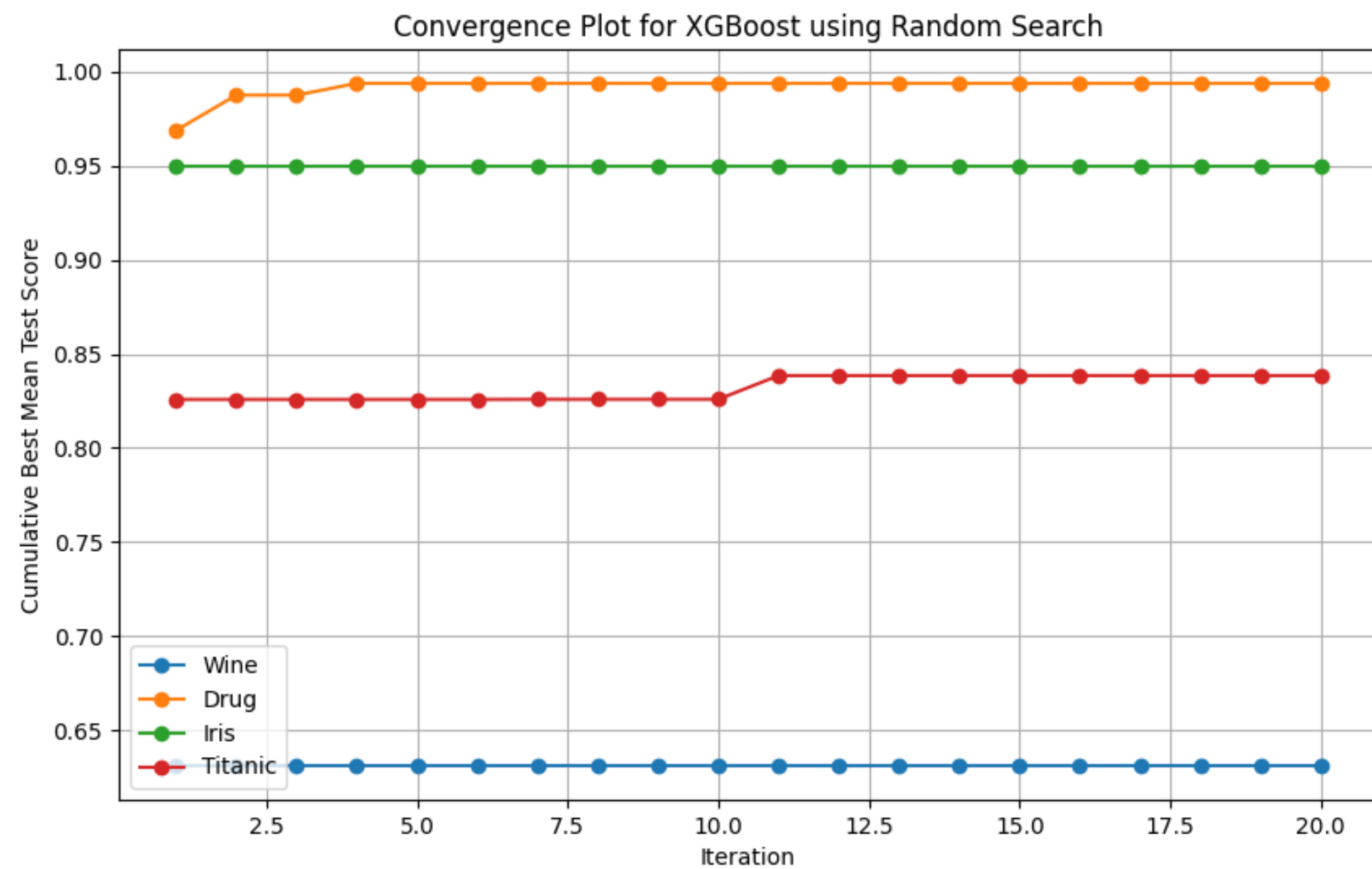
## Regresja logistyczna





# Wymanagna liczba iteracji

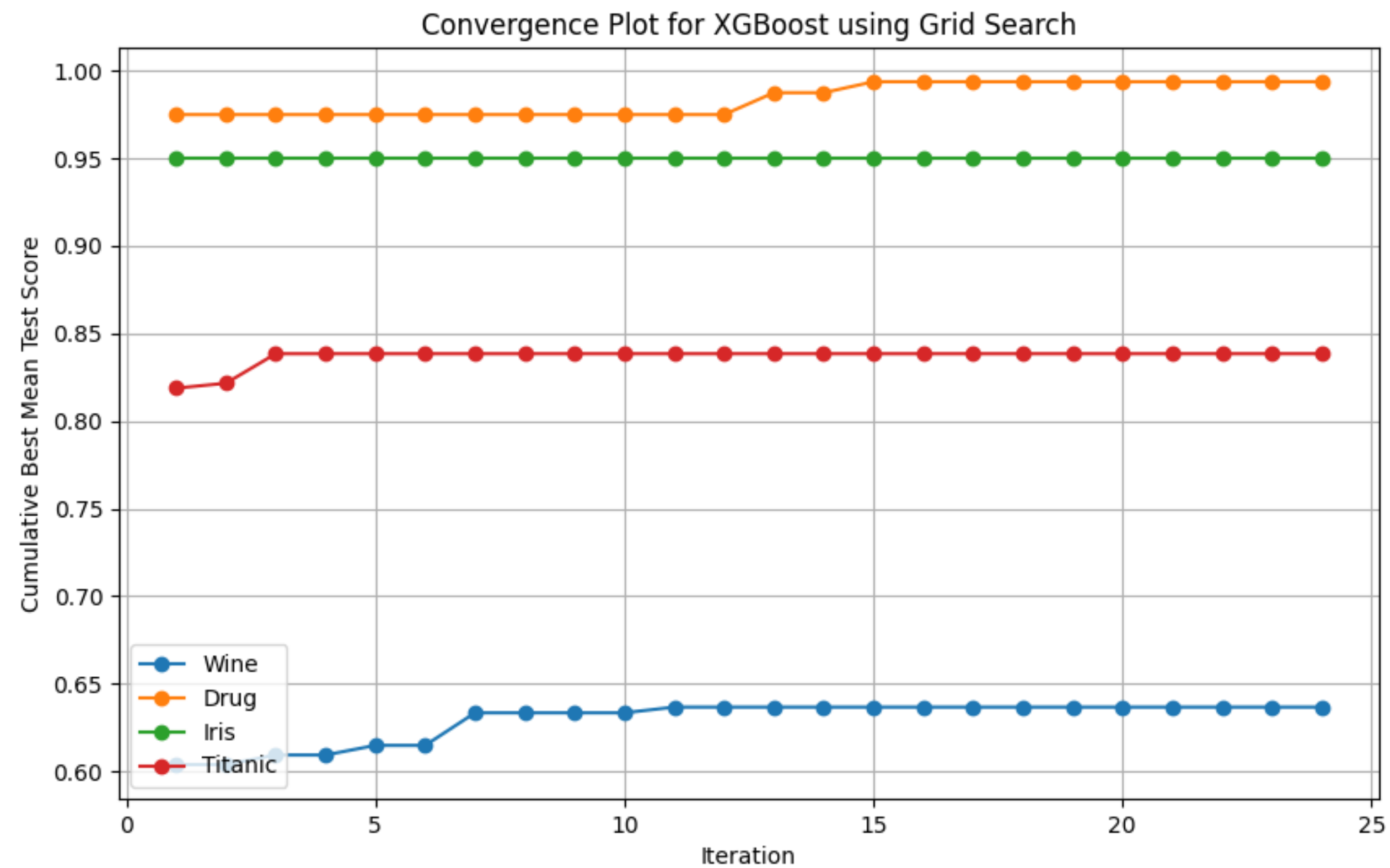
## XGBoost





# Wymanagna liczba iteracji

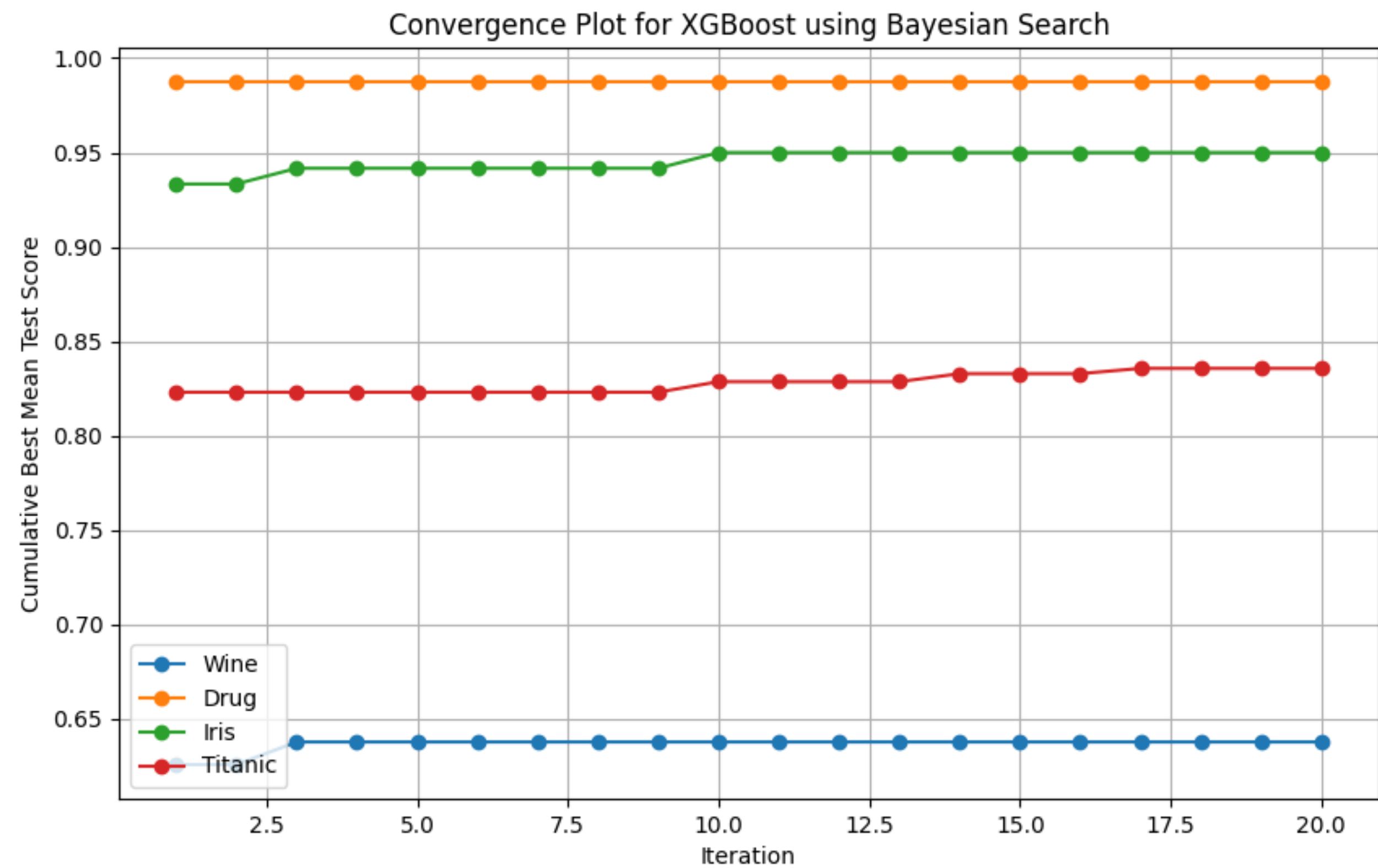
## XGBoost





# Wymanagna liczba iteracji

## XGBoost



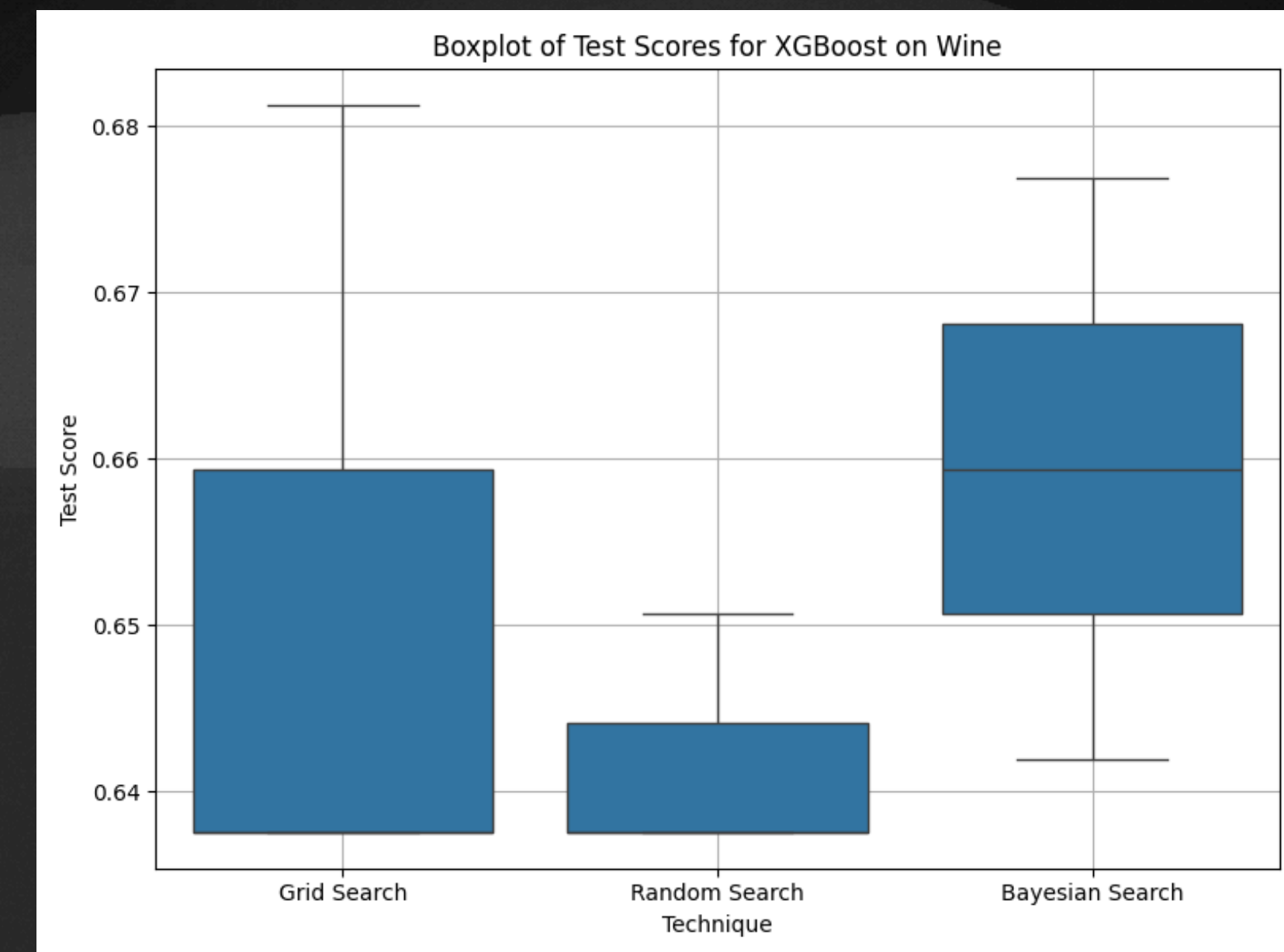
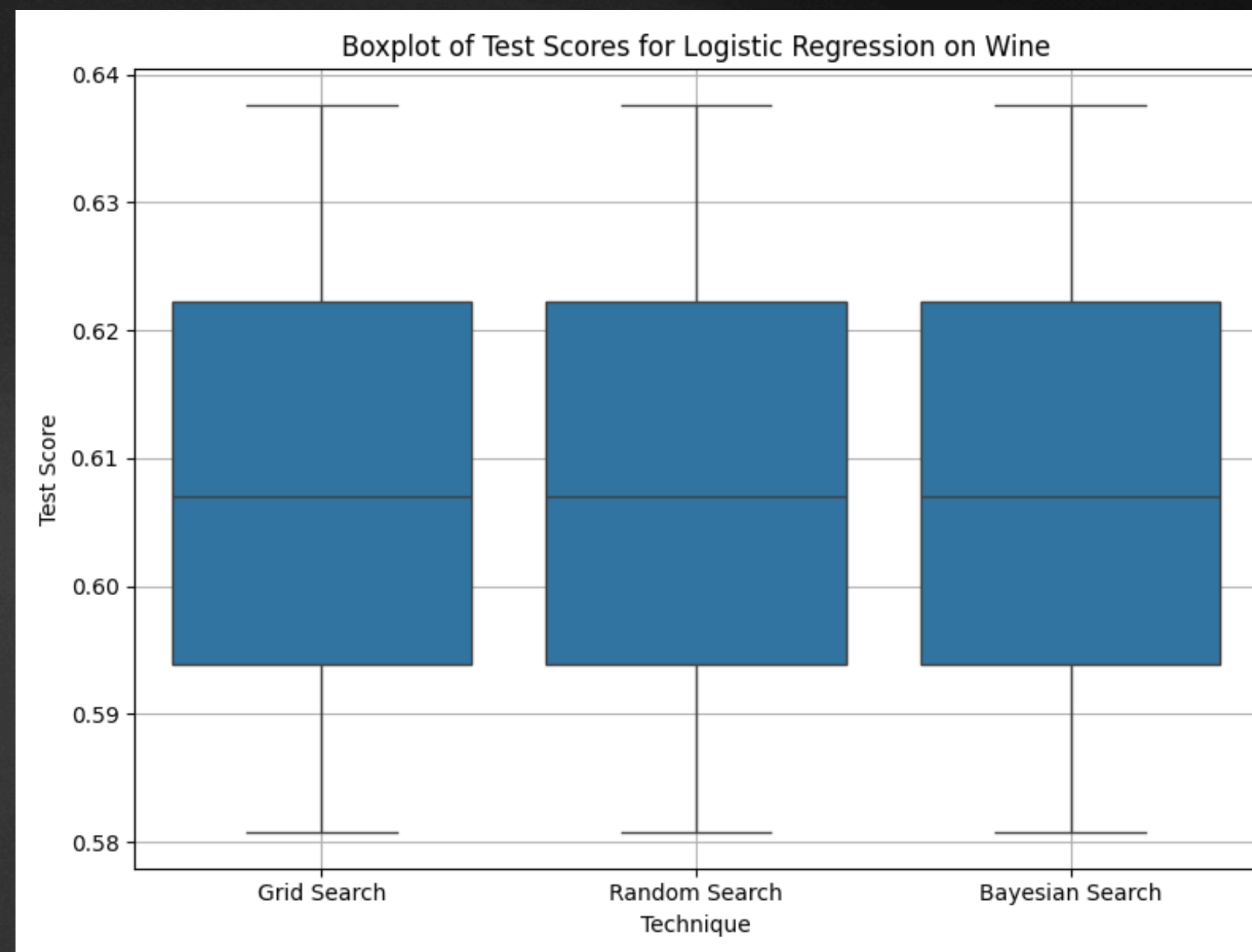
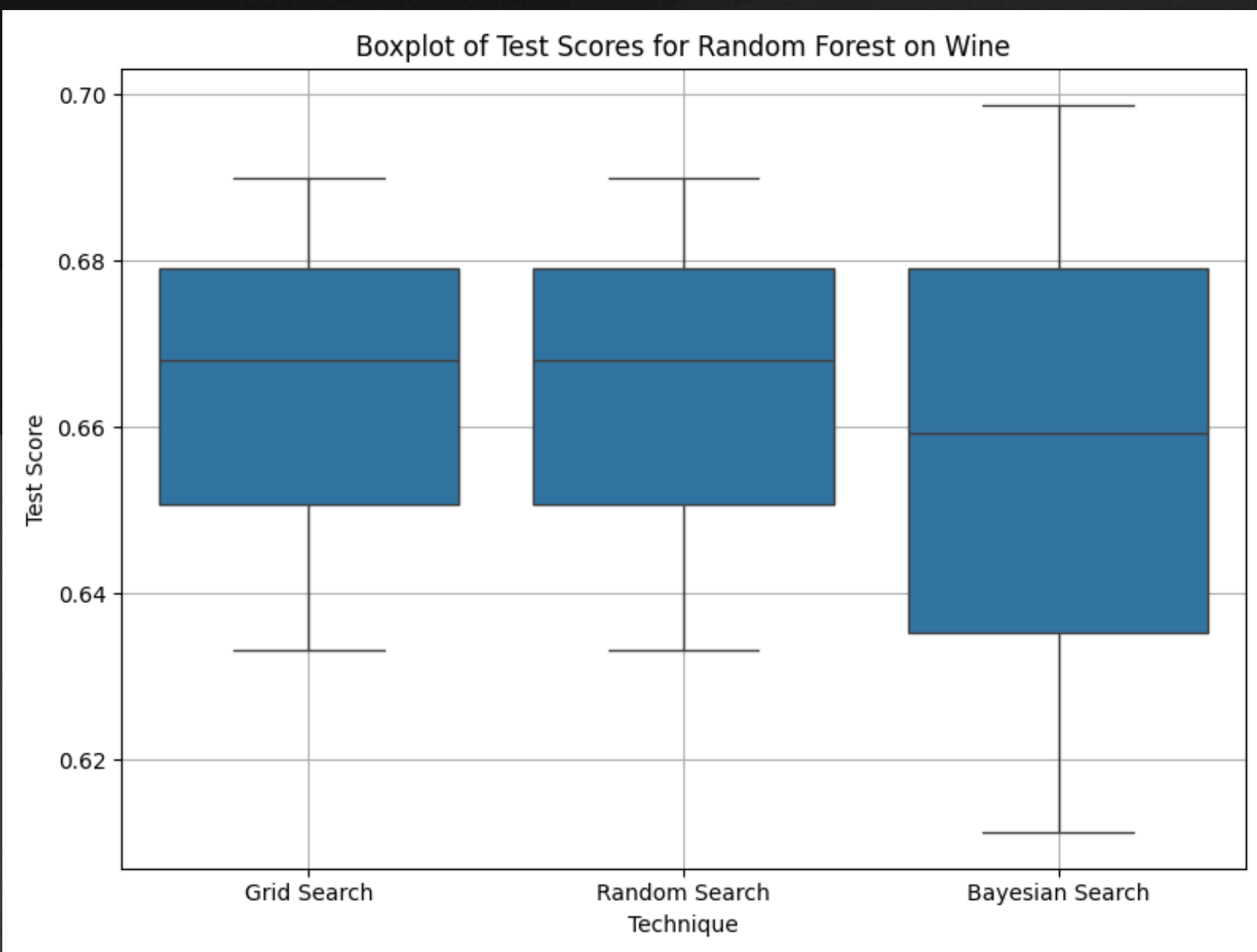


# Wizualizacja wyników



# Wizualizacja wyników

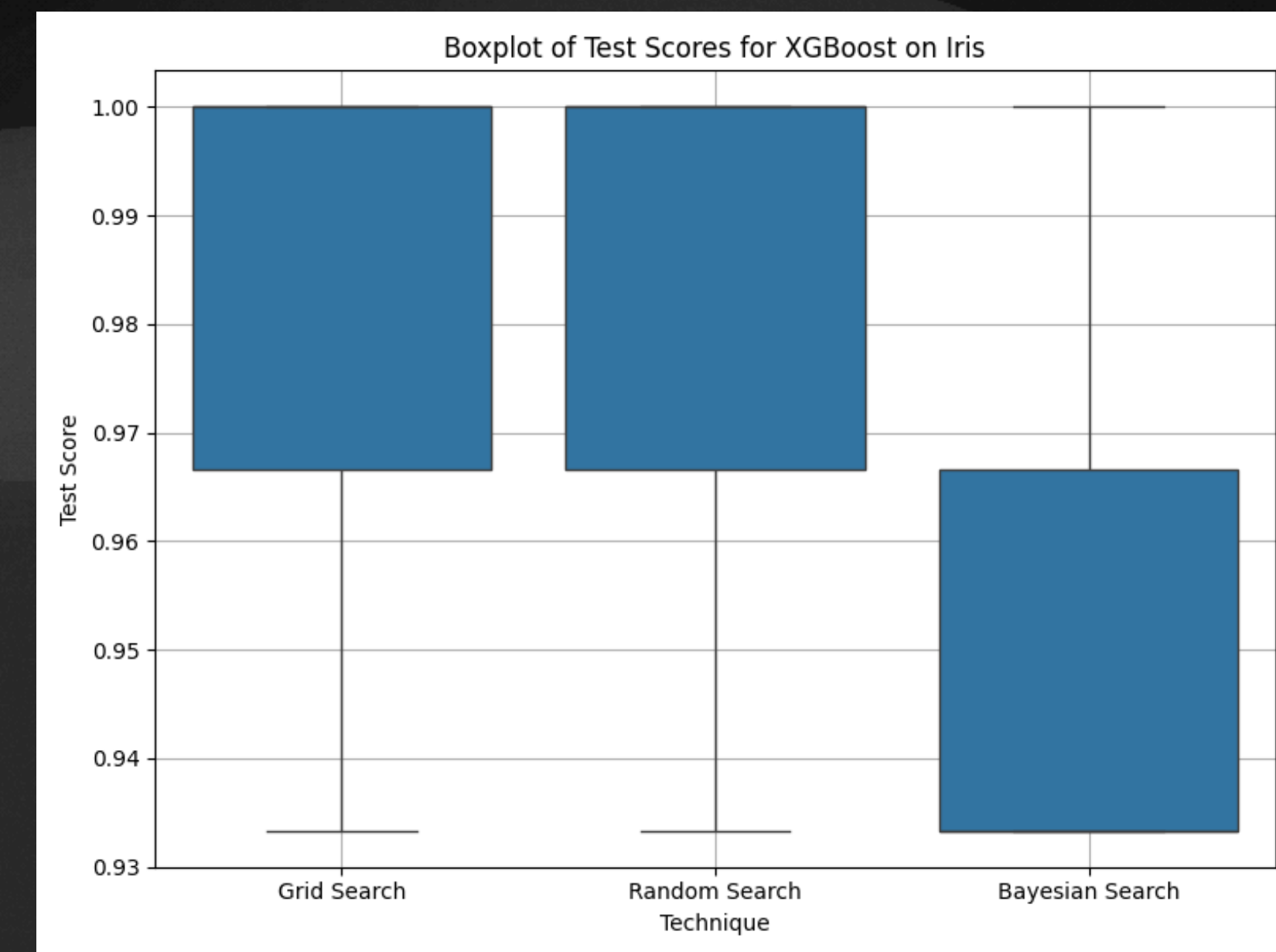
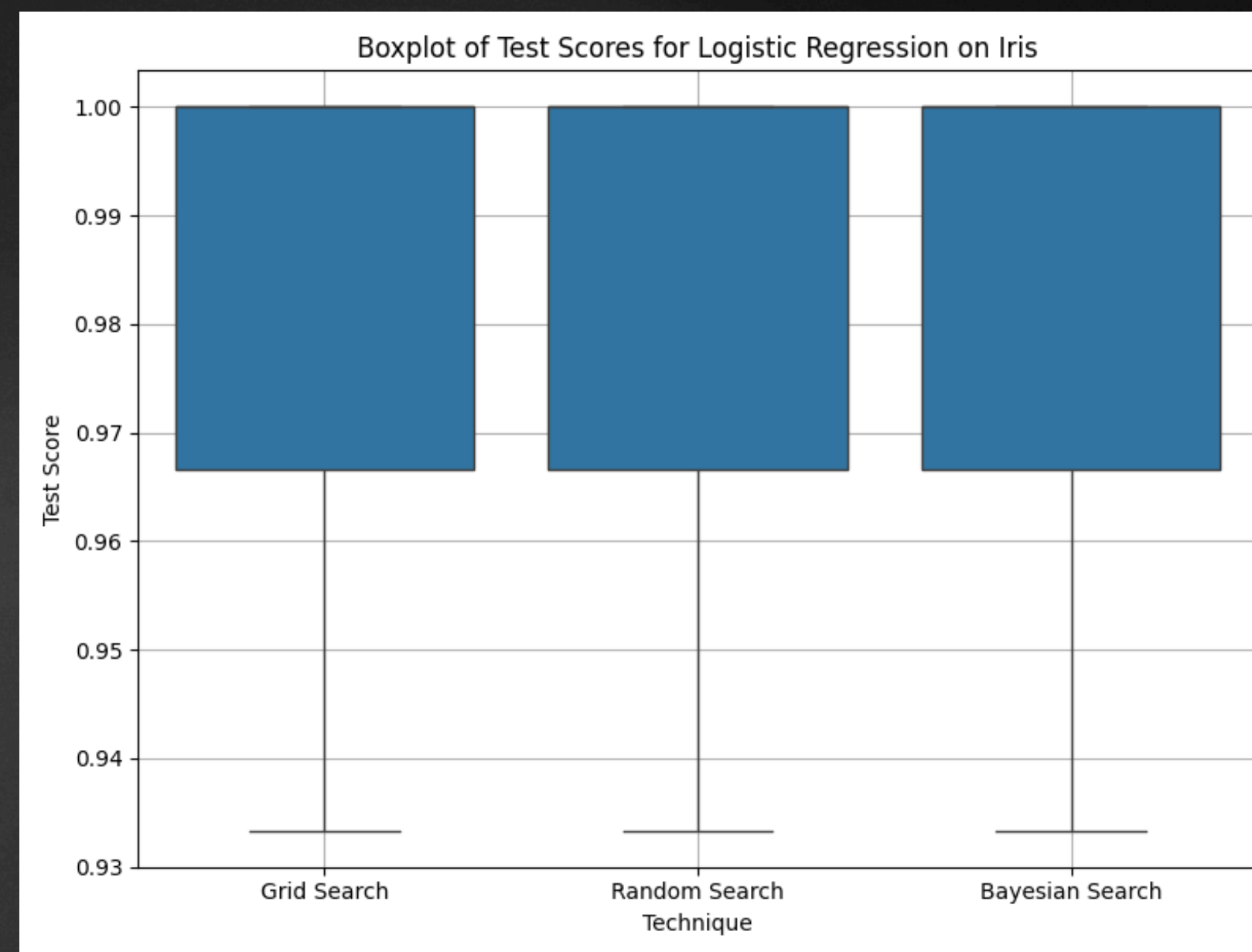
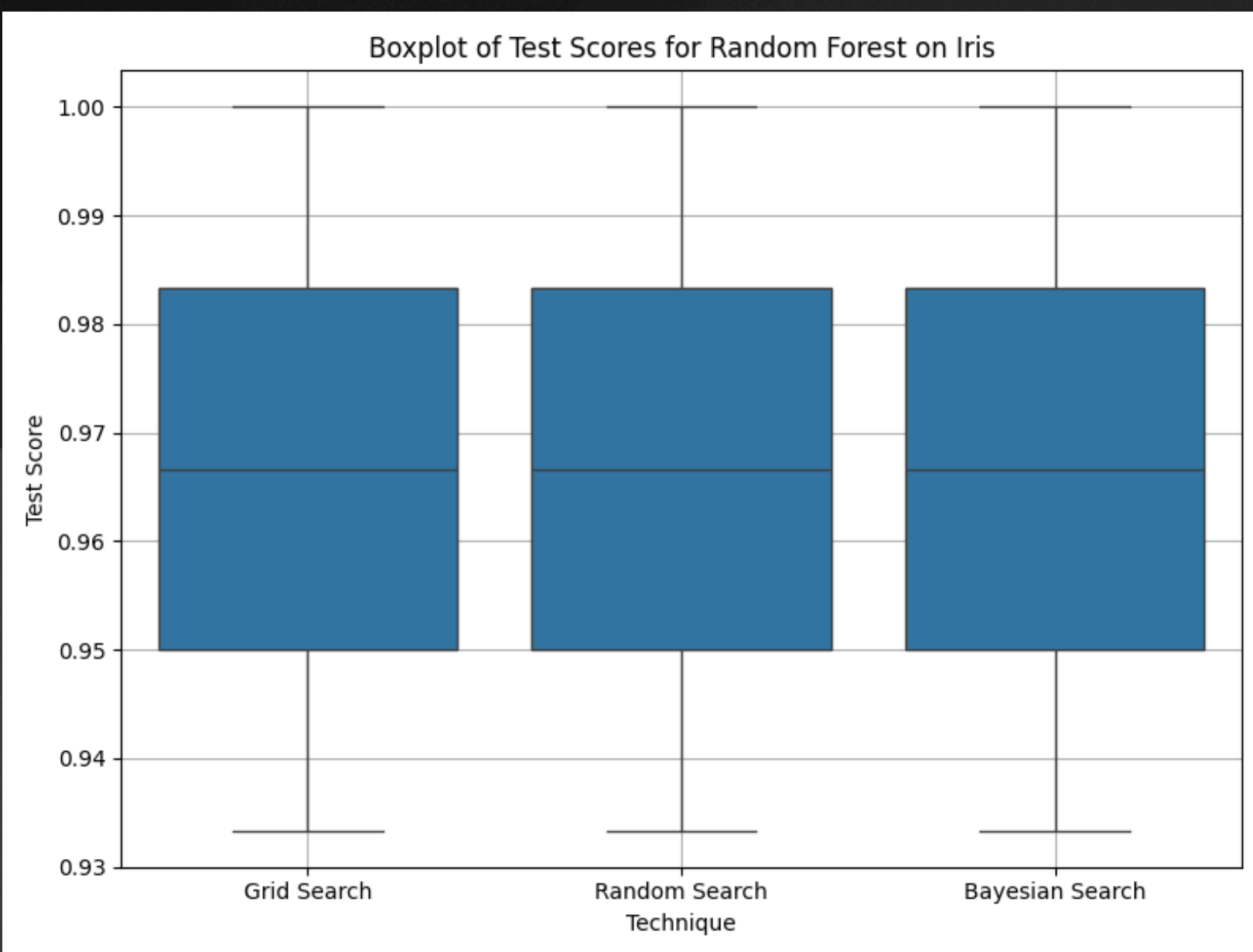
## Wyniki testów na zbiorze danych Wine





# Wizualizacja wyników

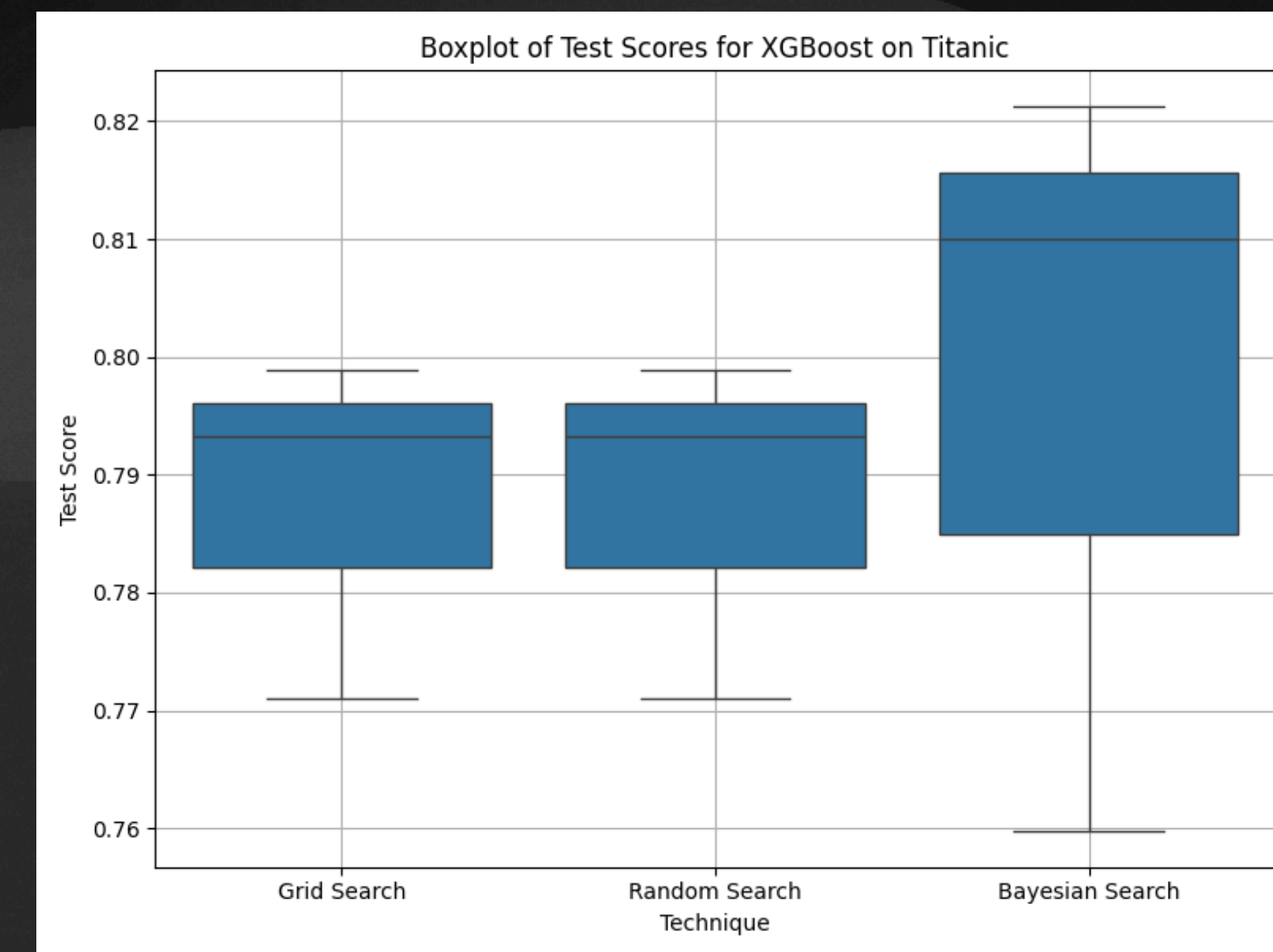
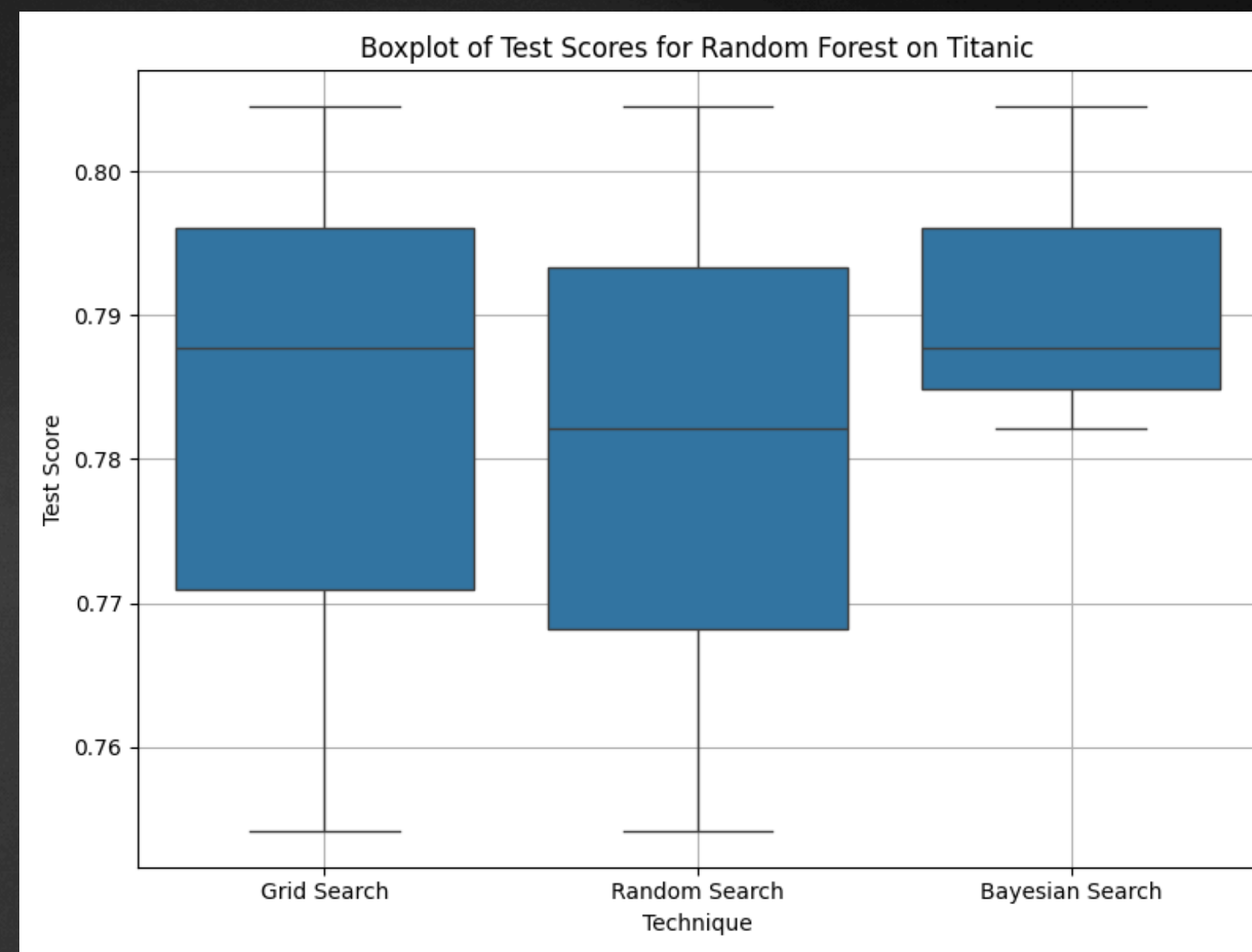
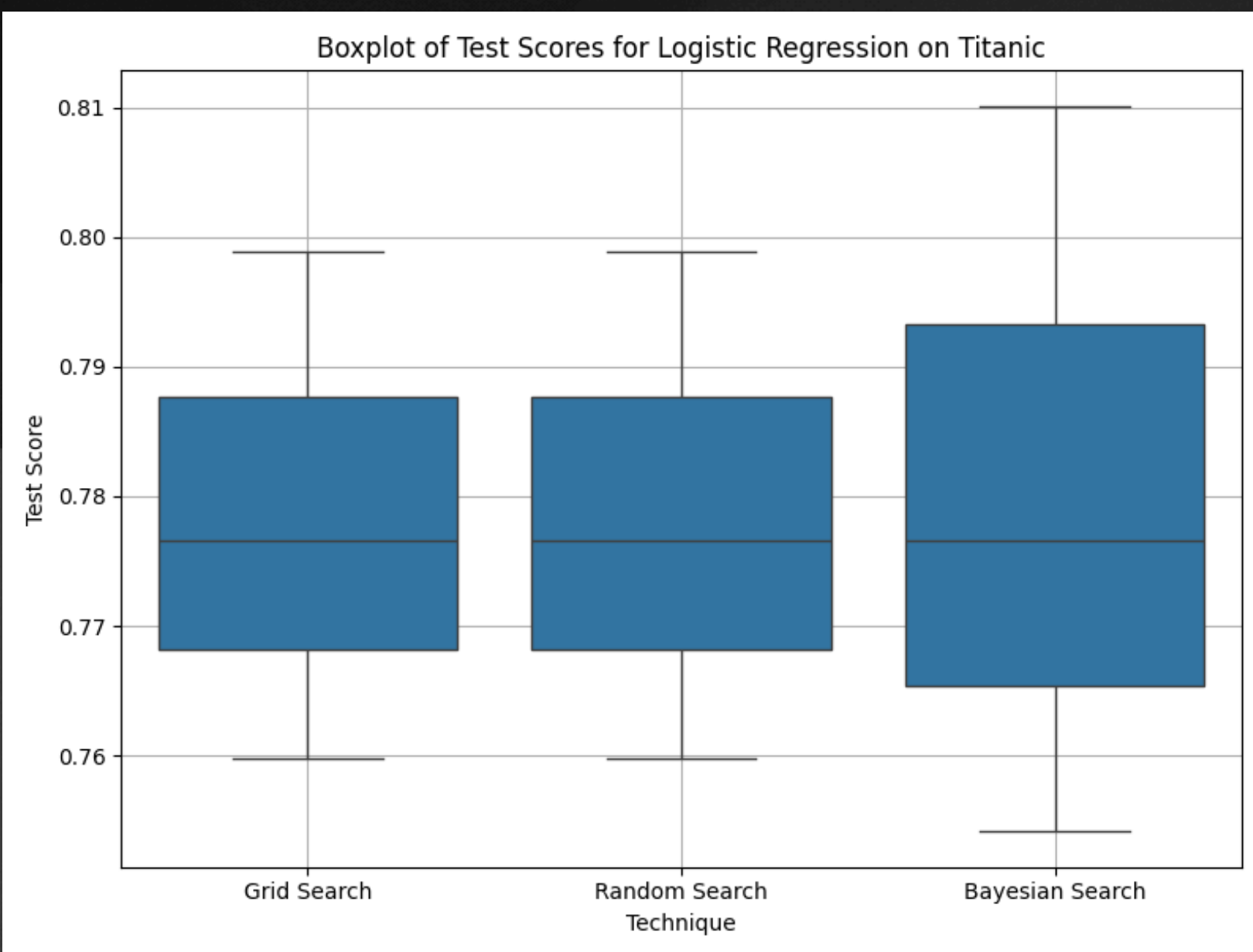
## Wyniki testów na zbiorze danych Iris





# Wizualizacja wyników

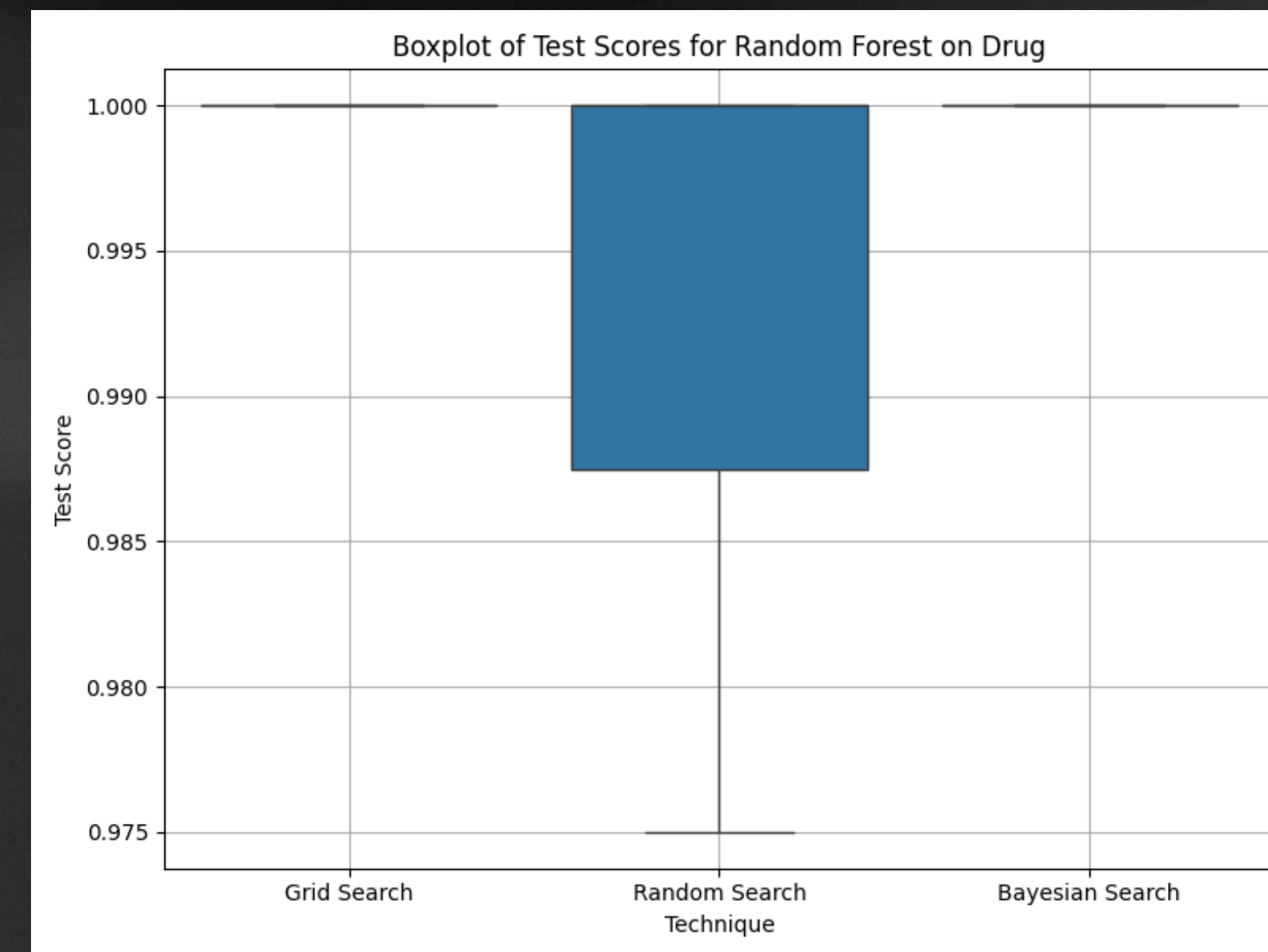
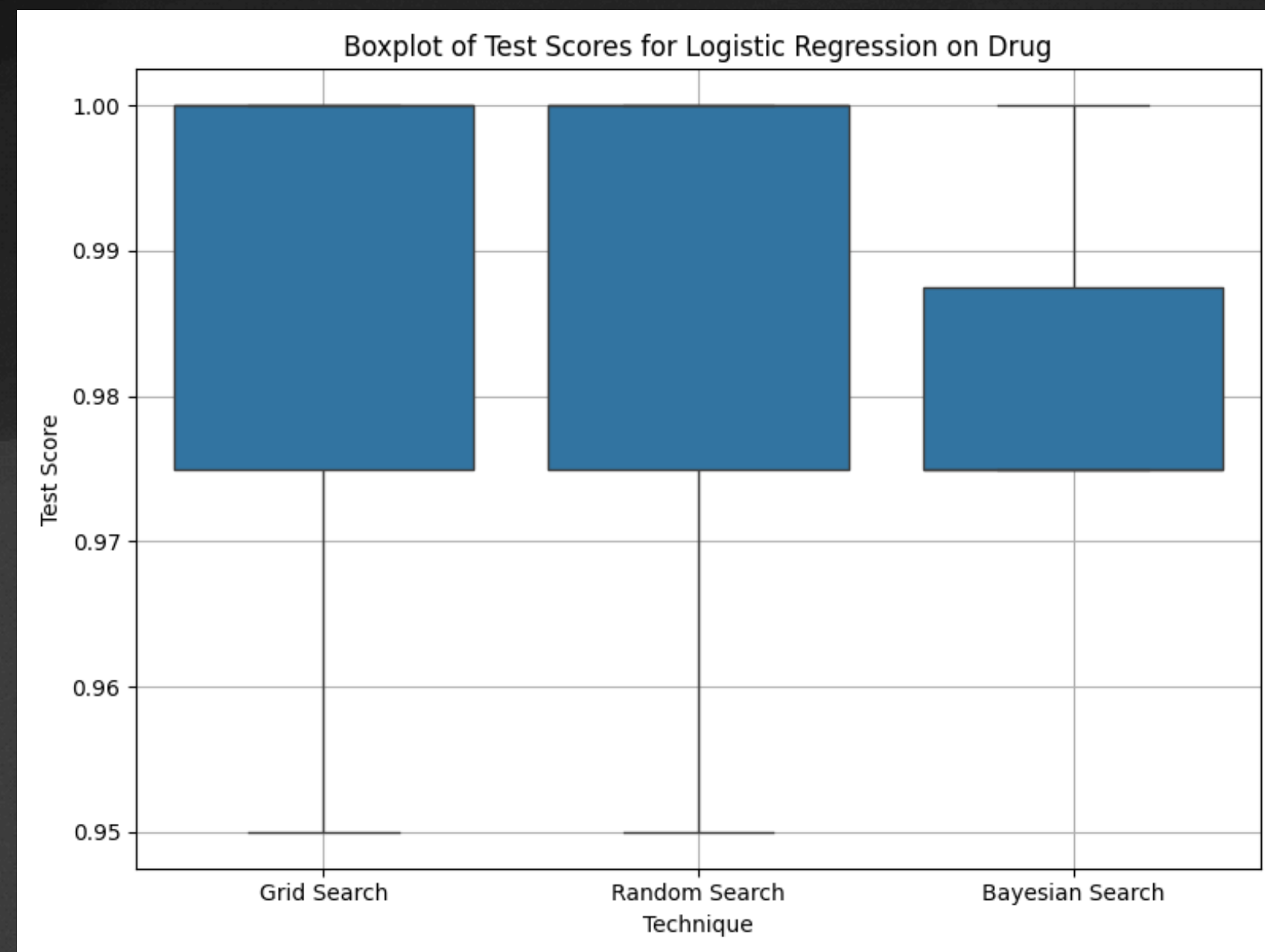
## Wyniki testów na zbiorze danych Titanic





# Wizualizacja wyników

## Wyniki testów na zbiorze danych Drug







Wnioski