Mataas Kaya Sweldo ko?

Using Optuna to tune Salary Grade Classification Models

ML1 Project Presentation - Google Slides

Dataset and Dictionary

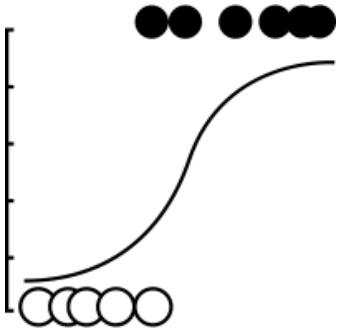
- The US Adult Census dataset is a repository of **48,842** entries extracted from the 1994 US Census database.
- Has **14** Features
- **Task:** Predict Salary Grade whether it is **>50K (1)** or **<= 50K (0)**

Dataset and Dictionary

Feature	Values
age	17-90 yrs old
workclass	Private, Local-gov, self-emp-inc, etc.
fnlwgt	Ex. 89814
education	Pre-school, HS-grad, Masters, Doctorate, etc.
marital-status	Never-married, *married, widowed, divorced, separated, etc.
occupation	Farming-Fishing, exec-managerial, tech-support, etc.

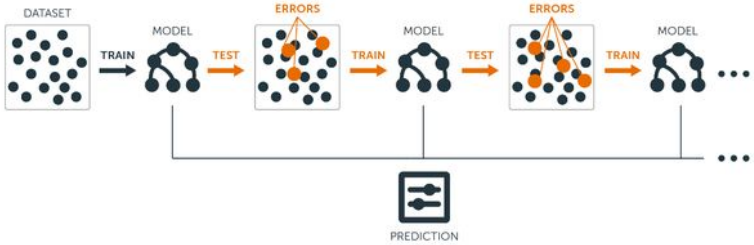
Feature	Values
relationship	Own-child, Husband, Unmarried, Wife, etc.
race	Black, White, Asian-Pac-Islander, etc.
sex	Male, Female
capital-gain	Ex. 7688
capital-loss	Ex. 7688
hours-per-week	Ex. 40 hrs
native-country	United-States, Philippines, Mexico, etc.

ML models Used



Logistic Regression

Random Forest



Gradient Boosting Classifier

“What’s the new learning here” You might ask 🤔 ?

OPTUNA

An open source hyperparameter optimization framework to automate hyperparameter search

DEMO

Basic Optuna Framework

```
def objective_function():
    <some code>
    </some code>
    return the thing you want to maximize or minimize

study = optuna.create_study(direction='maximize or minimize',
                            sampler=sampler of choice)
study.optimize(objective_function, n_trials=No.of Trials)
```

Basic Optuna Framework

```
1  def objective_function():
2      <some code>
3      </some code>
4      return the thing you want to maximize or minimize
5
6  study = optuna.create_study(direction='maximize or minimize',
7                              sampler=sampler of choice)
8  study.optimize(objective_function, n_trials=No.of Trials)
```

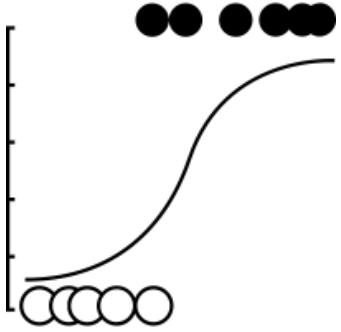
Basic Optuna Framework

```
1  def objective_function():
2      <some code>
3      </some code>
4      return the thing you want to maximize or minimize
5
6  study = optuna.create_study(direction='maximize or minimize',
7                              sampler=sampler of choice)
8  study.optimize(objective_function, n_trials=No.of Trials)
```

Code Time!

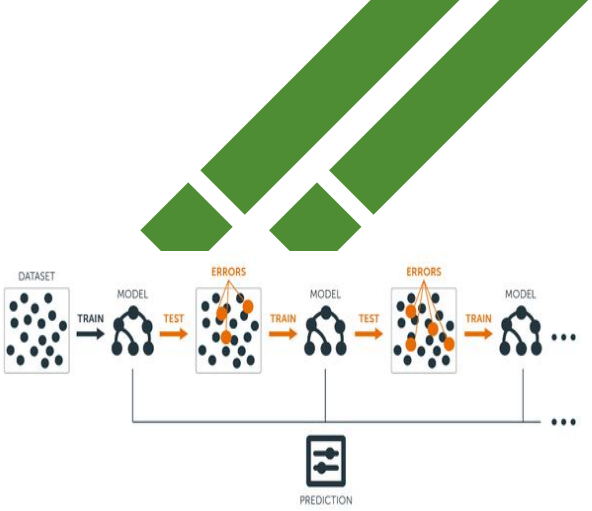
ML models Summary

PCC	62.72 %
GOAL	78.40 %



Logistic Regression

Random Forest

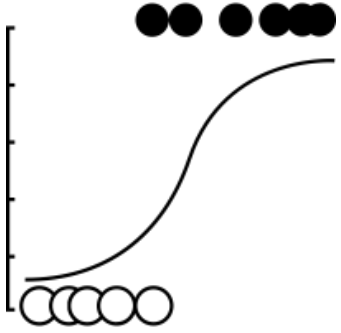


Gradient Boosting Classifier

	Logistic Regression	Random Forest	Gradient Boosting Classifier
Best Accuracy	84.86%	85.78%	86.51%
Best Parameters	Penalty : L1, C : 0.69039	n_estimators : 370, max_depth : 10, Max_features : 'sqrt'	n_estimators: 500, max_depth: 5, max_features: 'sqrt', learning_rate: 0.085
Top 2 Predictors	Capital gain (+) > Education (+)	Married to a civilian > Capital Gain	Married to a civilian > Capital Gain

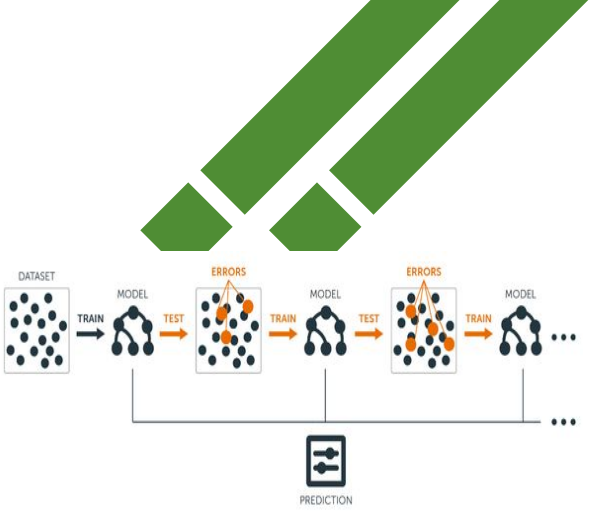
ML models Summary

PCC	62.72 %
GOAL	78.40 %



Logistic Regression

Random Forest

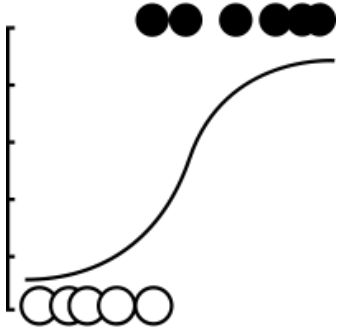


Gradient Boosting Classifier

Best Accuracy	84.86%	85.78%	86.51%
Best Parameters	Penalty : L1, C : 0.69039	n_estimators : 370, max_depth : 10, Max_features : 'sqrt'	n_estimators: 500, max_depth: 5, max_features: 'sqrt', learning_rate: 0.085
Top 2 Predictors	Capital gain (+) > Education (+)	Married to a civilian > Capital Gain	Married to a civilian > Capital Gain

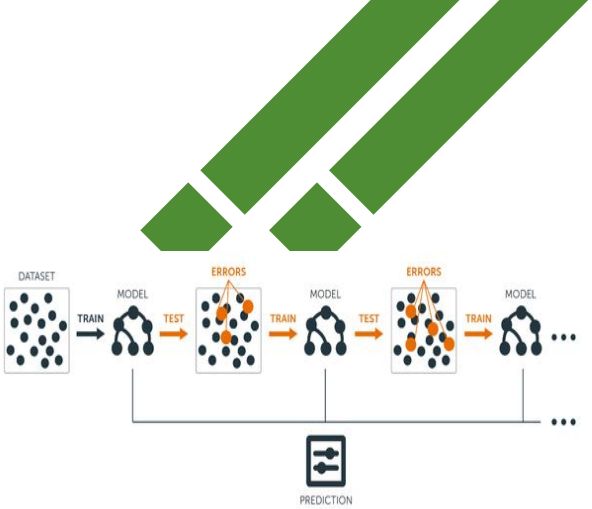
ML models Summary

PCC	62.72 %
GOAL	78.40 %



Logistic Regression

Random Forest

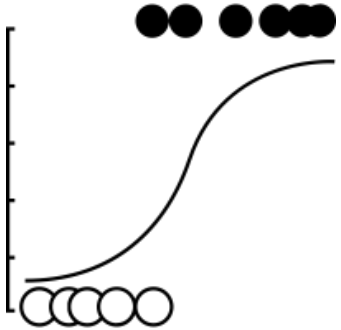


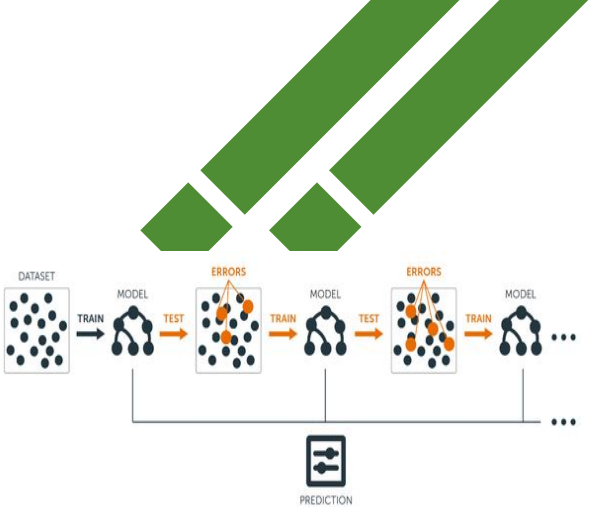
Gradient Boosting Classifier

	Logistic Regression	Random Forest	Gradient Boosting Classifier
Best Accuracy	84.86%	85.78%	86.51%
Best Parameters	Penalty : L1, C : 0.69039	n_estimators : 370, max_depth : 10, Max_features : 'sqrt'	n_estimators: 500, max_depth: 5, max_features: 'sqrt', learning_rate: 0.085
Top 2 Predictors	Capital gain (+) > Education (+)	Married to a civilian > Capital Gain	Married to a civilian > Capital Gain

ML models Summary

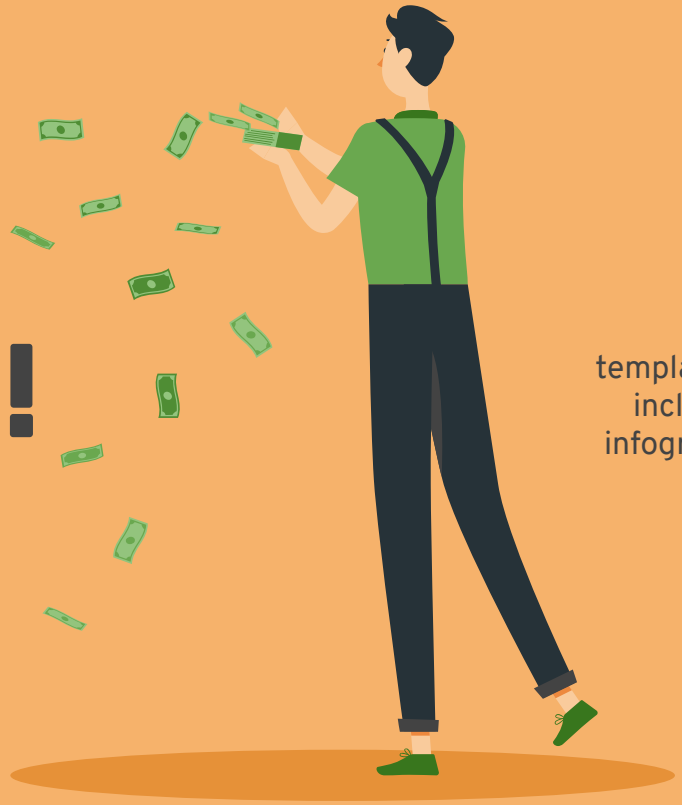
PCC	62.72 %
GOAL	78.40 %





	Logistic Regression	Random Forest	Gradient Boosting Classifier
Best Accuracy	84.86%	85.78%	86.51%
Best Parameters	Penalty : L1, C : 0.69039	n_estimators : 370, max_depth : 10, Max_features : 'sqrt'	n_estimators: 500, max_depth: 5, max_features: 'sqrt', learning_rate: 0.085
Top 2 Predictors	Capital gain (+) > Education (+)	Married to a civilian > Capital Gain	Married to a civilian > Capital Gain

Thanks!



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.