Premium Session

Data Science Project

Best Practices & Tips

KIRTHIKKA DEVI VENKATARAM

Lead at WWC Data Science

Session Agenda

- Advanced evaluation methods and metric selection
- Improving our project - advanced techniques
- Implementing our project
- Takeaways & Perks

Machine Learning Problems

- Supervised

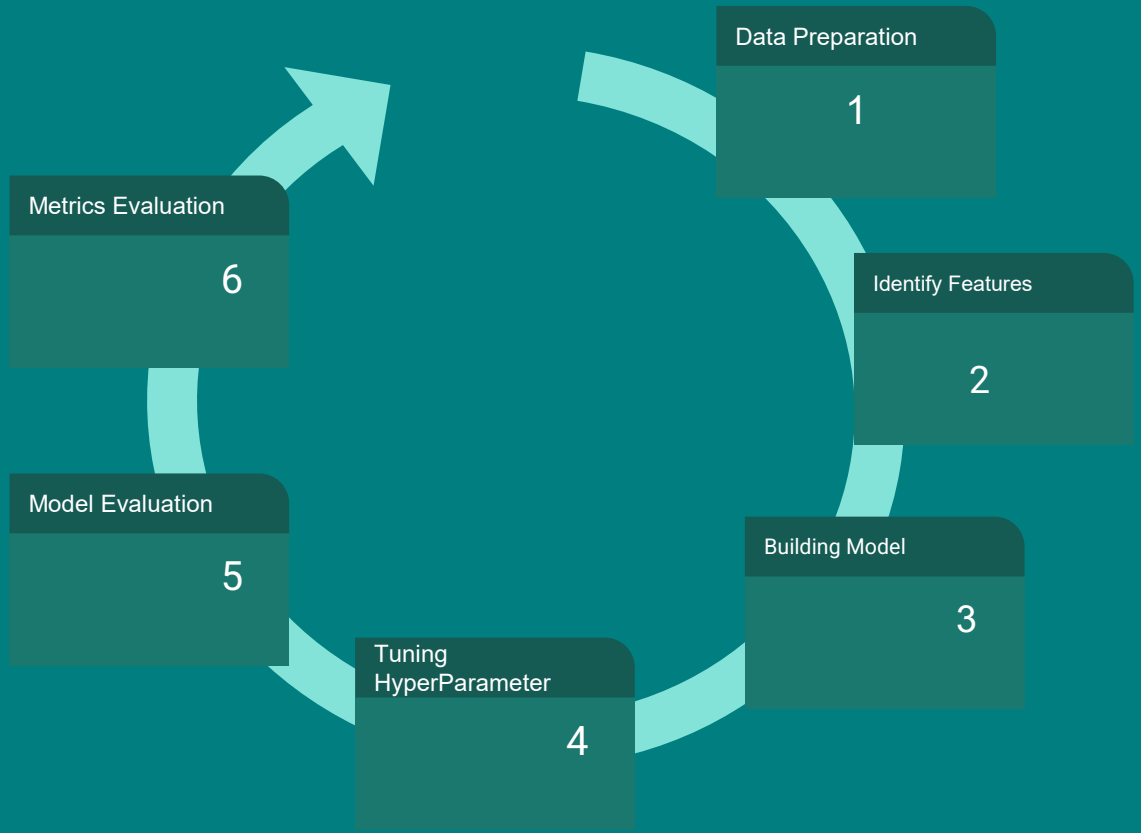
- Unsupervised

- Semi supervised

- Reinforcement

WOMEN WHO CODE® /data-science

Data Science Life Cycle



WOMEN WHO CODE®
/data-science

Types of Classification Problems

Binary Classification

- Balanced Dataset Binary Classification
- Imbalanced Dataset Binary Classification

Multiclass Classification

ML Classification Problem

- **Logistic Regression**
- **K-Nearest Neighbors**
- **Decision Trees**
- **Support Vector Machines (SVM)**
- **Naive Bayes**

ML Classification Problem - KPI & Metrics

- Accuracy
- Precision
- Recall
- F1 Score
- ROC
- AUC
- Entropy
- Gini Index
- Information Gain

Python: Libraries such as Scikit-Learn, TensorFlow, and PyTorch.

Types of Other Regression Problems

- Linear Regression
- Polynomial Regression
- Ridge Regression
- Lasso Regression

ML Regression Problem

- **Linear Regression**
- **Polynomial Regression**
- **Support Vector Regression**
- **Decision Tree Regression**
- **Random Forest Regression**

ML Regression Problem - KPI & Metrics

- Mean Absolute Error
- Mean Squared Error
- Root Mean Squared Error
- Coefficient of Determination

Python: Libraries such as Scikit-Learn, TensorFlow, and PyTorch.

ML Models

- Linear Regression
- Logistics Regression
- Decision Tree
- Random Forest
- Neural Networks

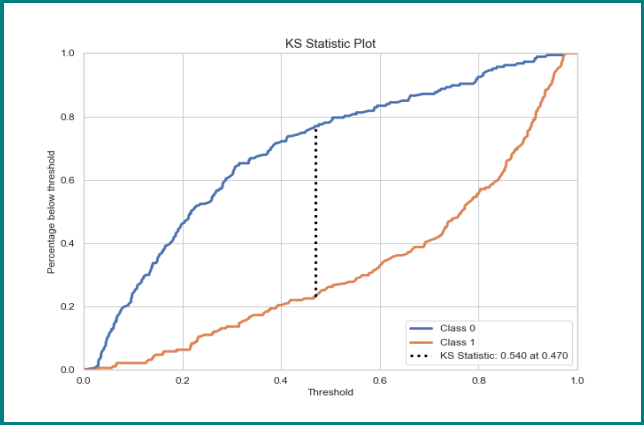
ML Model Evaluation & Validation methods

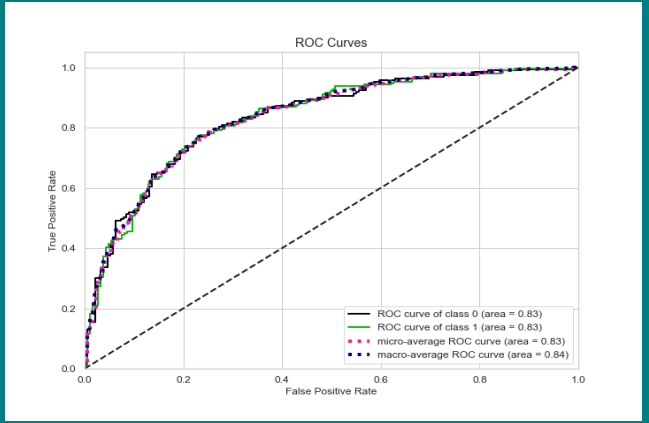
- Hold-out
- K-folds
- Leave-one-out
- Leave-p-out
- Stratified K-folds
- Repeated K-folds
- Nested K-folds
- Time series CV

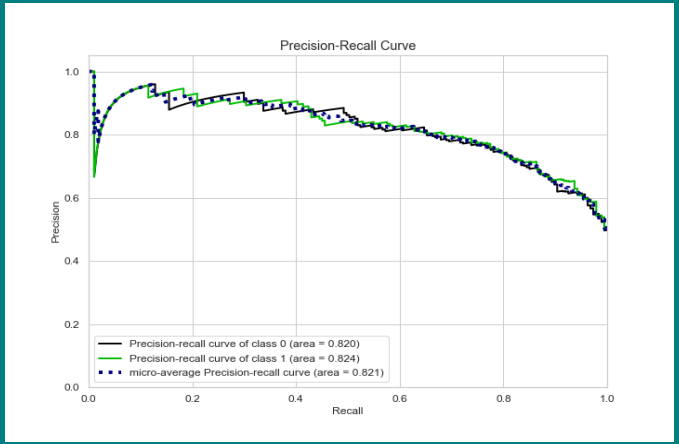
KPI & Metrics - Machine Learning Problem

Link to General Evaluation Metrics

KPI identification

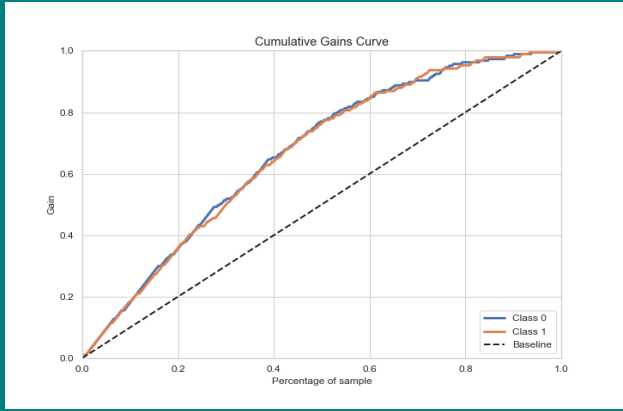


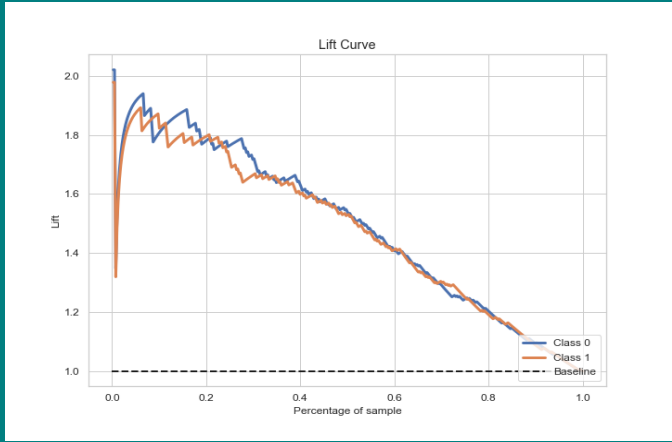




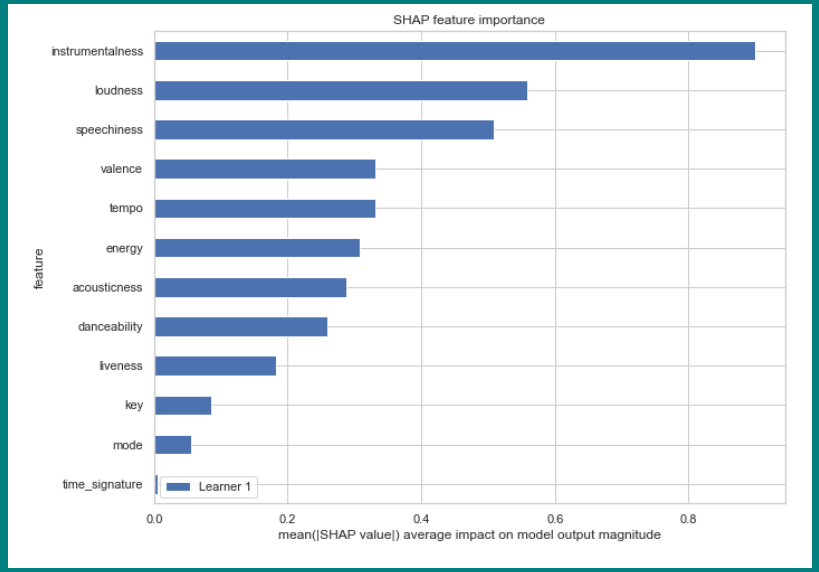
WOMEN WHO CODE® /data-science

KPI identification

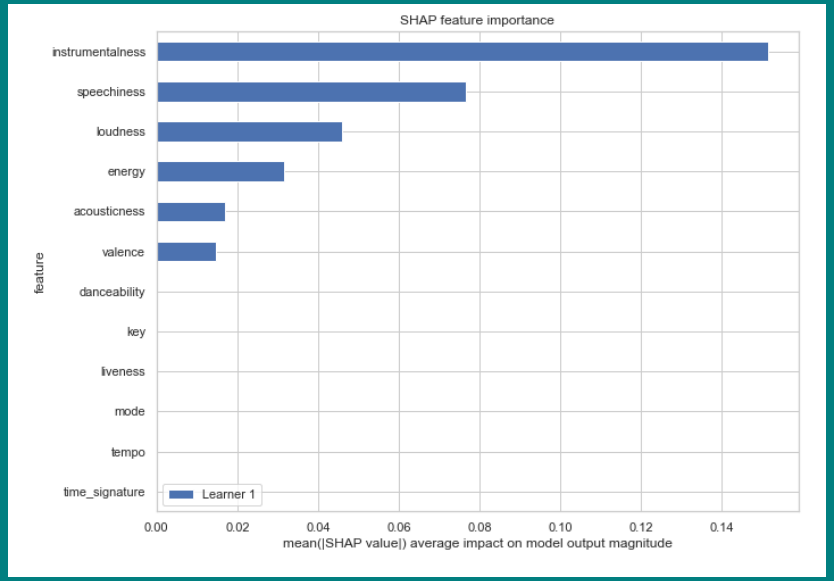




SHAP - SHapley Additive exPlanations



SHAP feature importance for XGBoost



SHAP feature importance for Decision Tree

WOMEN WHO CODE® /data-science

AutoML Demo

Takeaways

Thank you!

WOMEN WHO CODE®
/data-science