



ORACLE

Oracle AutoMLx

Giulia Carocari

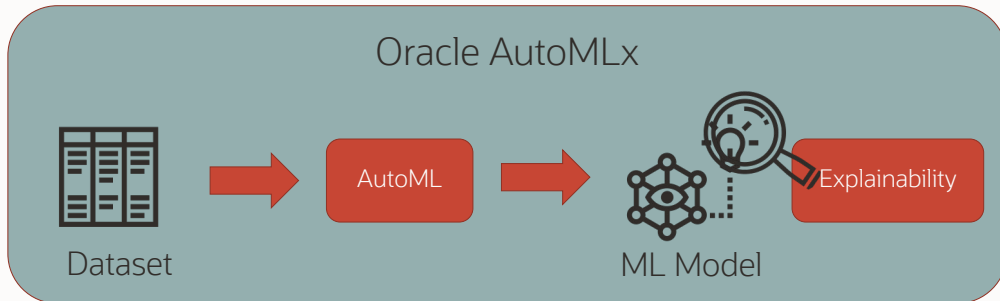
Member of Technical Staff, Oracle Labs Zurich

The data scientist pipeline



Creating a model:

- Which model?
- Are my features good (enough)?
- What hyper-parameter configuration?



Using a model:

- Can I trust my model?
- Is my model “fair”?
- Does it meet regulatory requirements?



AutoMLx

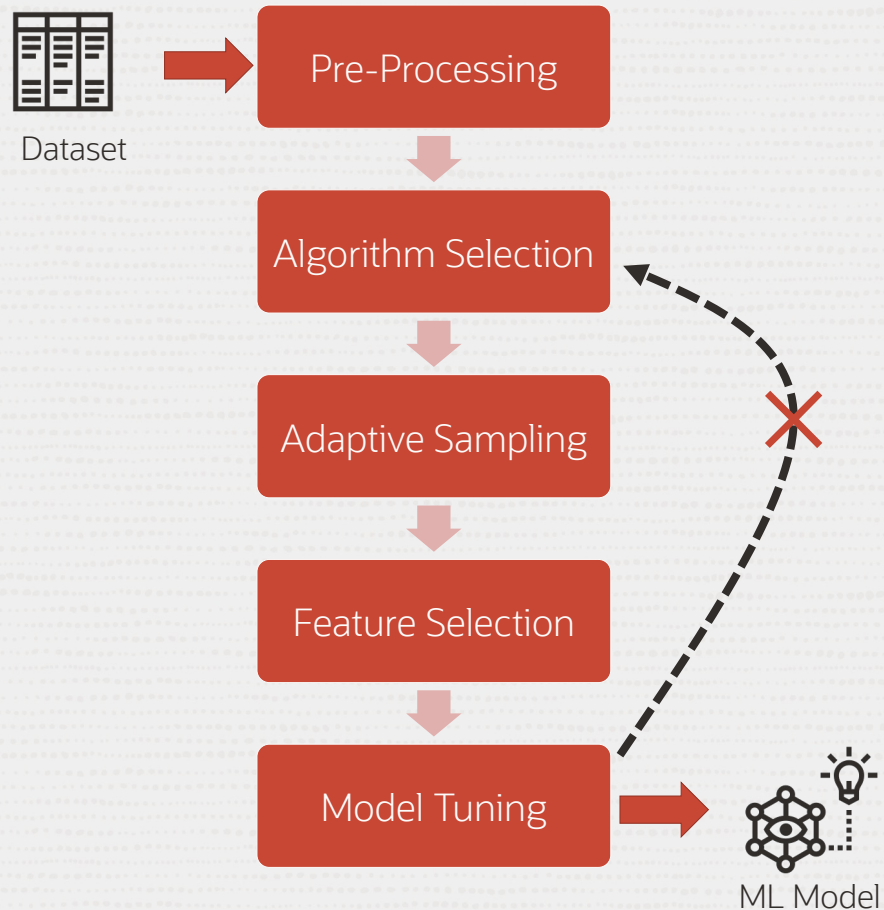
```
from automl import Pipeline
```

```
# 'regression', 'forecasting' and 'anomaly detection' also supported  
pipeline = Pipeline(task='classification')
```

```
pipeline.fit(X, y)
```

```
y_pred = pipeline.predict(X_test)
```


Oracle's AutoML pipeline



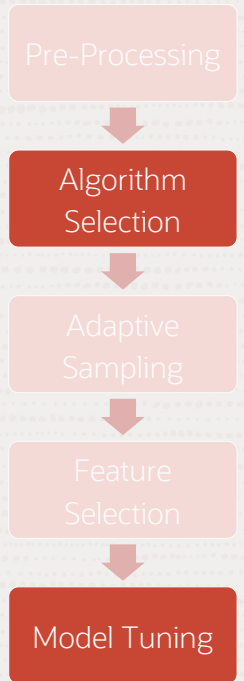
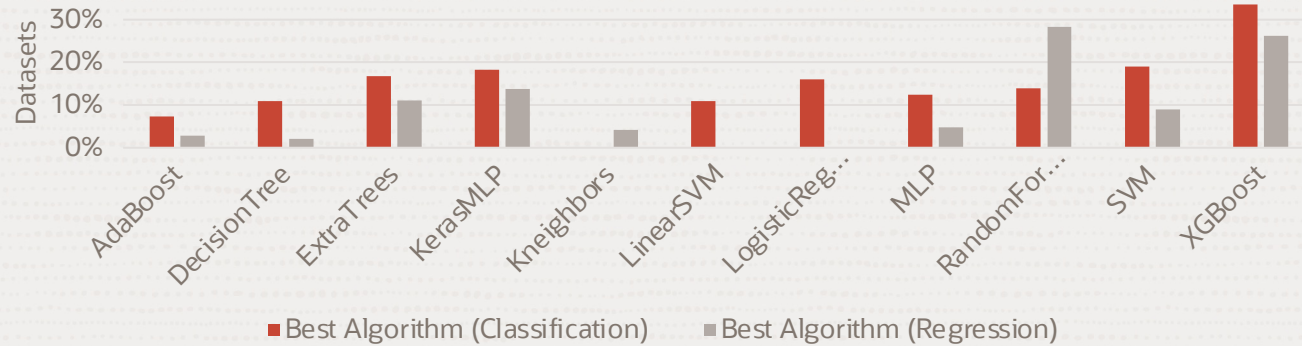
Traditional AutoML uses:

- Combined algorithm selection and hyper-parameter configuration

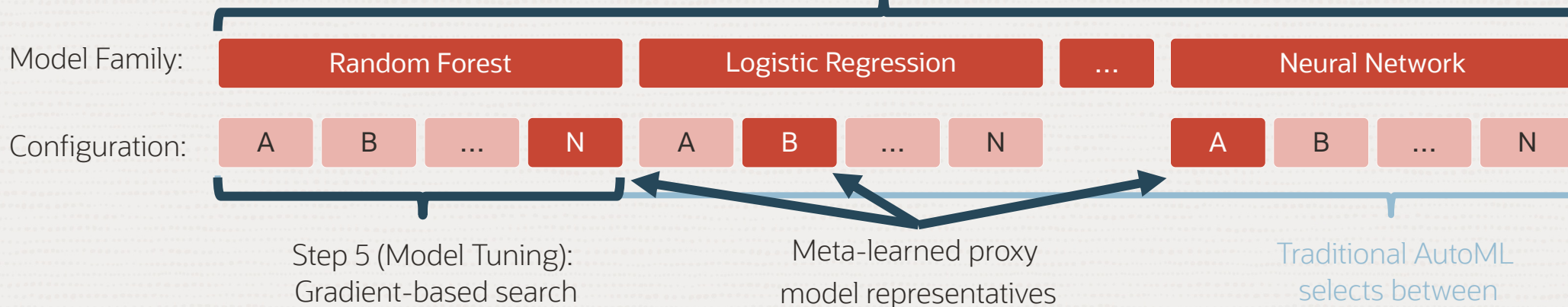
Our secret sauce?

- We never look back!

Algorithm selection & model tuning



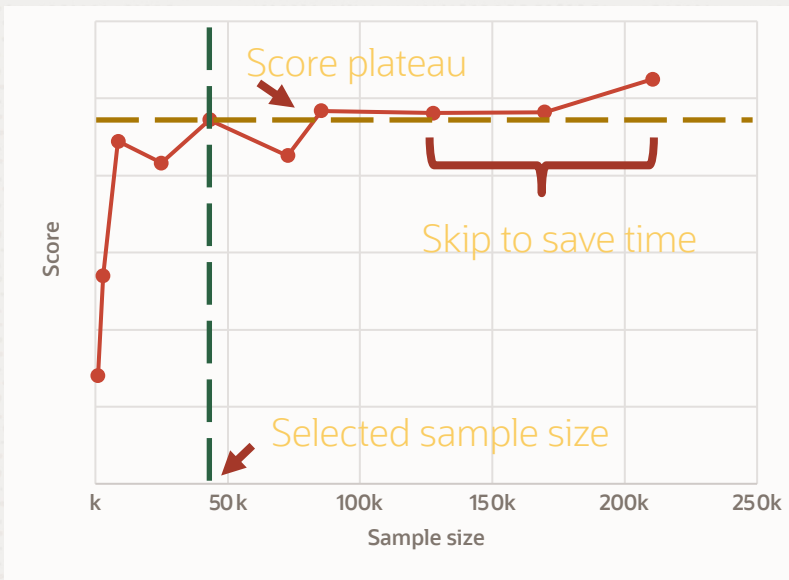
Step 2 (Algorithm Selection):
We select between



Adaptive data reduction

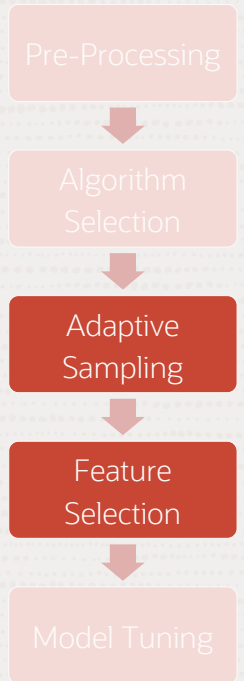
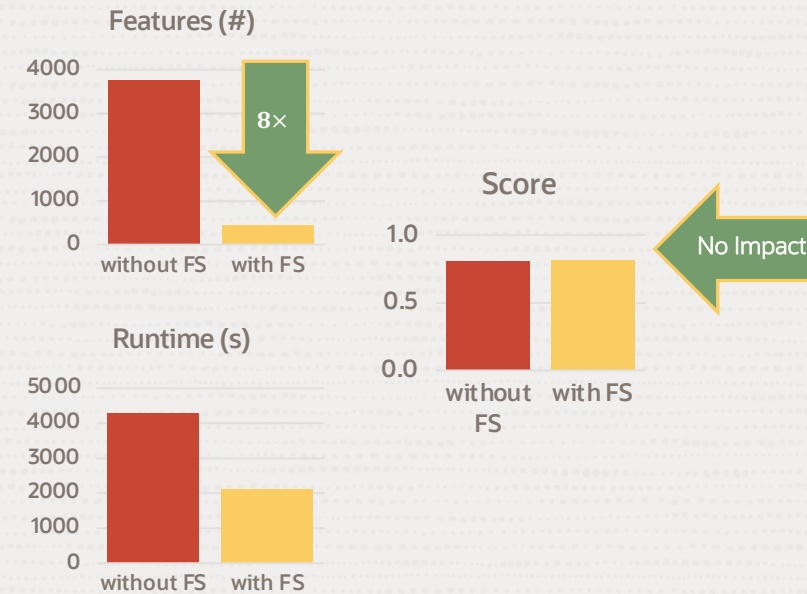
Adaptive Sampling

- Subsample rows for faster training
- Speeds up model search

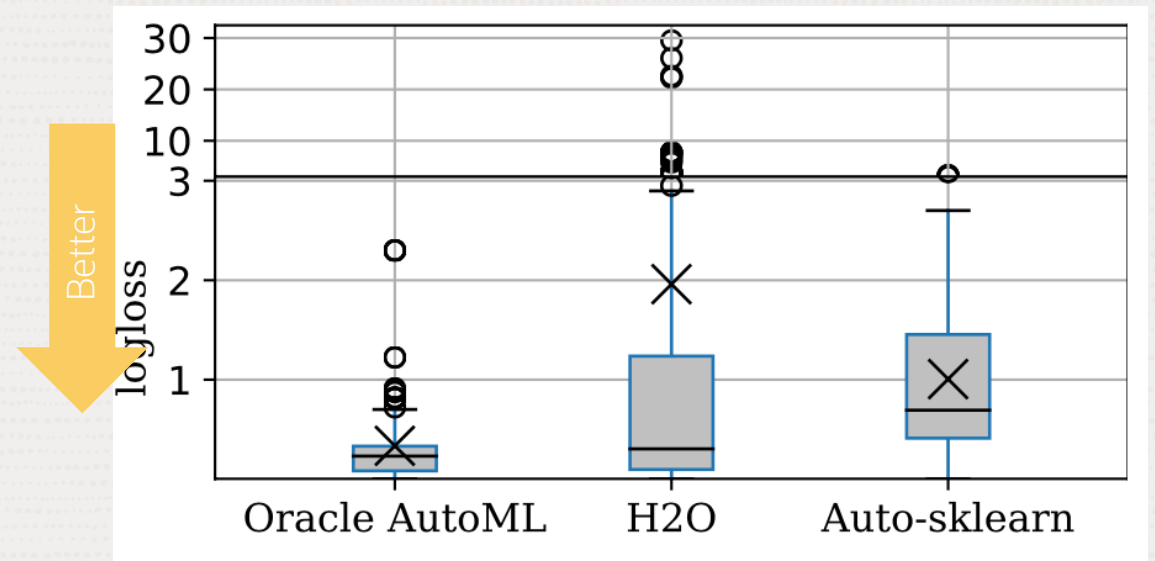
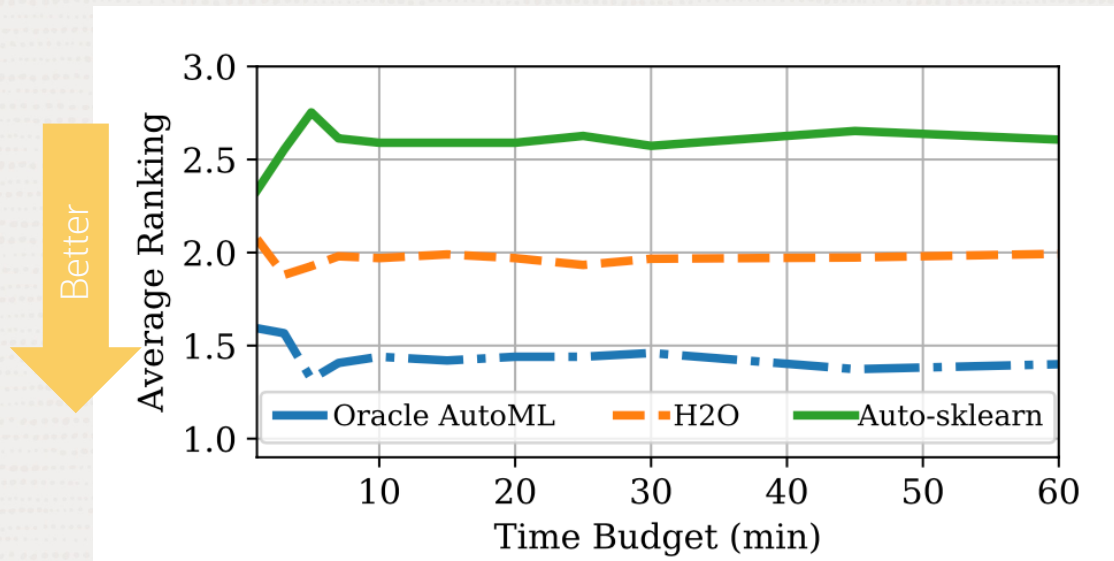


Feature Selection

- Subsample columns for faster training
- Can also reduce overfitting



Oracle AutoML Benchmarking



3.5 – 4x faster
and better scores

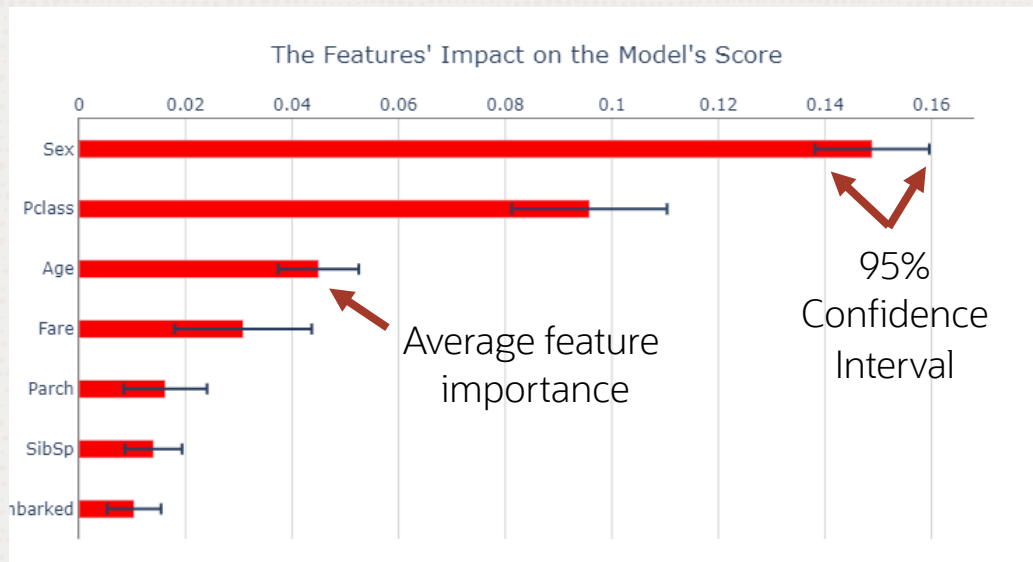
[Yakovlev, Anatoly, et al. "Oracle automl: a fast and predictive automl pipeline." *Proceedings of the VLDB Endowment* 13.12 \(2020\): 3166-3180.](#)



AutoMLx

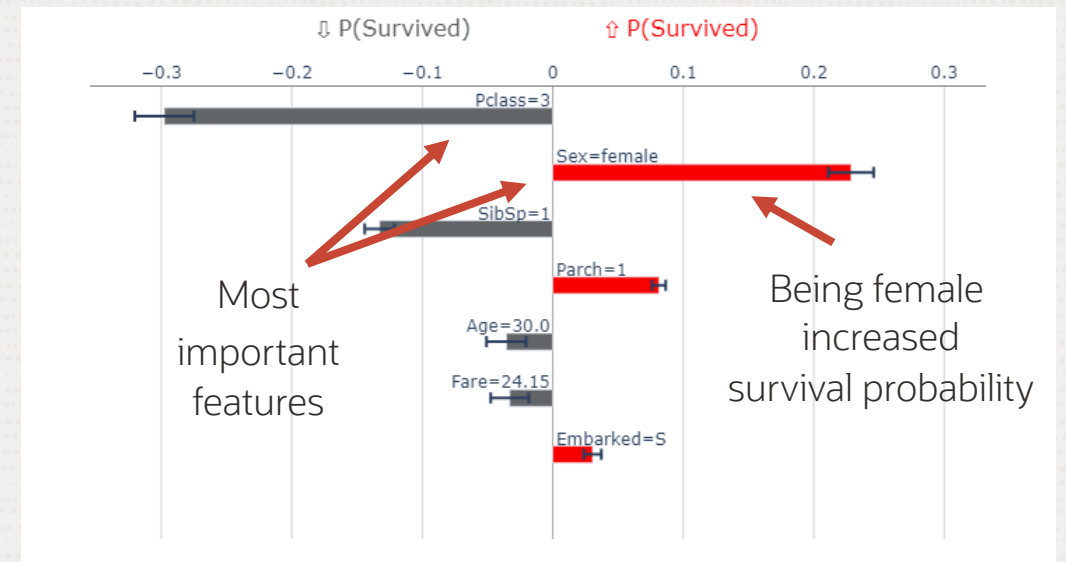
```
from automl import MLEExplainer
# Can be an AutoML pipeline or scikit-learn model
explainer = MLEExplainer(model, X, y, task)
# Global feature importance
explainer.explain_model()
# Local feature importance
explainer.explain_prediction(X_test)
# Partial dependence plot
explainer.explain_feature_dependence(feature)
```


Feature importance examples – titanic dataset



Global (model) feature importance

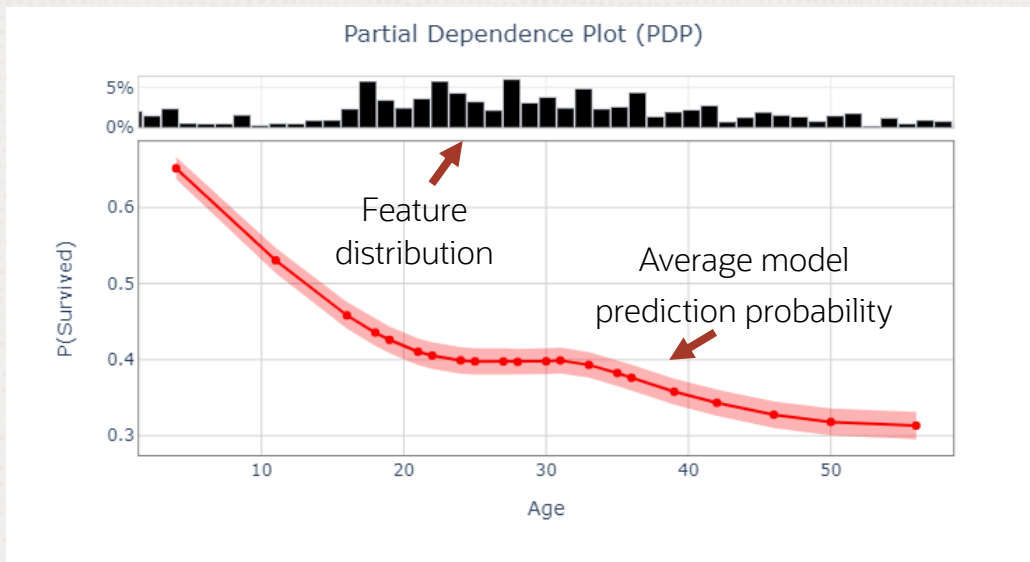
`explainer.explain_model()`



Local (prediction) feature importance

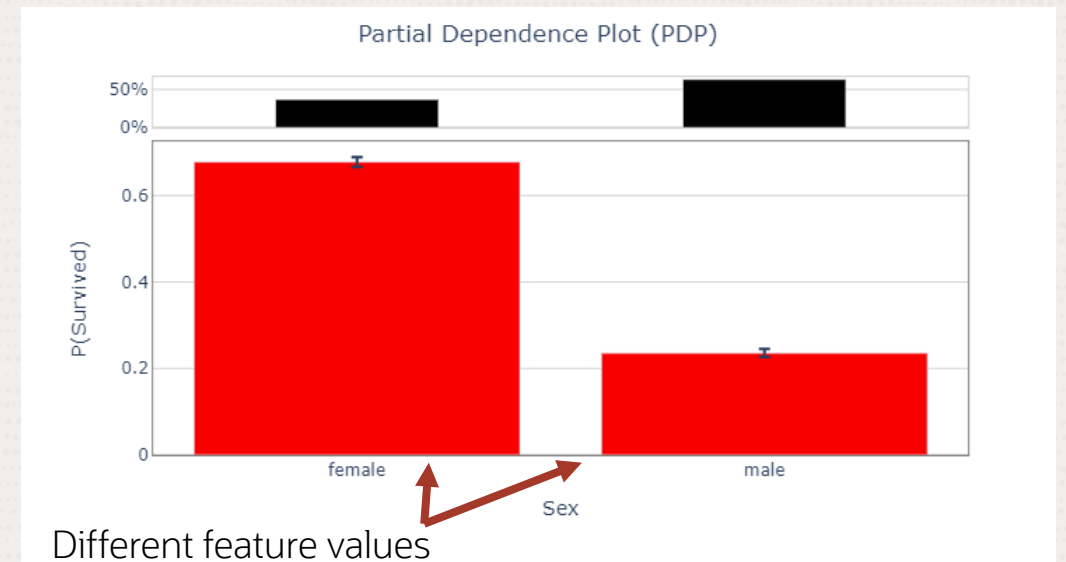
`explainer.explain_prediction(X_test)`

Feature dependence examples – titanic dataset



Continuous feature PDP

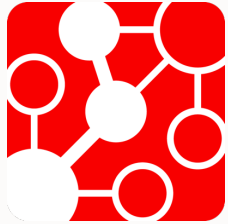
`explainer.explain_feature_dependence('age')`



Categorical feature PDP

`explainer.explain_feature_dependence('sex')`

Machine learning projects at Oracle Labs



Graph machine learning
with PGX



ML for healthcare



AutoMLx



ML for cybersecurity



Recommender systems

Oracle Labs internship page: <https://labs.oracle.com/> > Internships

