Institute of Actuaries of India
Statutory body under an Act of Parliament

Expanding the horizon, strengthening the core
20th Global Conference of Actuaries
4th - 6th March, 2019 | Mumbai, India

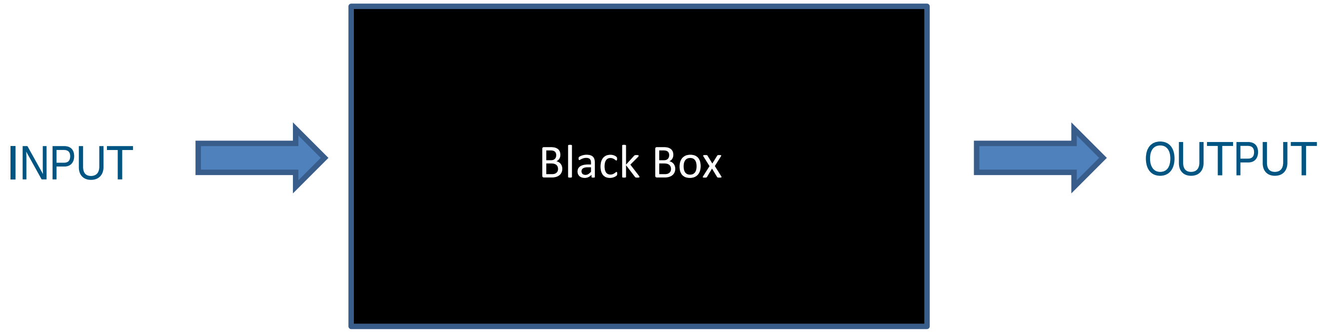
Being Transparent About Our Machine Learning Predictions

Speaker(s):
Hemant Kumar (FIAI)
Abhijit Kulkarni (PhD)

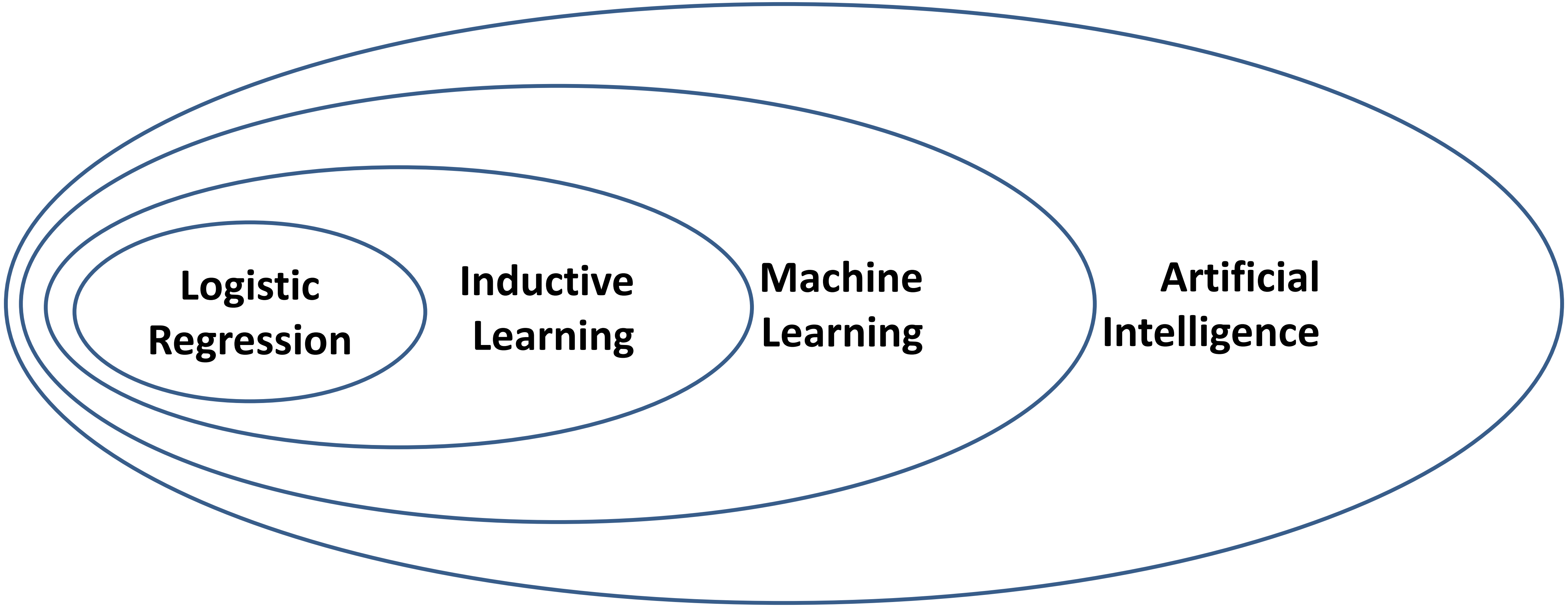
Session # P8
Dated: March 06, 2019

Need for Explainability of Predictive Models

- Actuarial work place – Redefined
- Data Analytics – Extensive use across industries. And this is just the beginning.
- There is a need to remove barriers to make its use more reliable and widespread
- Models should be explainable – Stakeholders want it. Regulators mandate it.
- General Data Protection Regulation (GDPR) requirements
- Need to explain the Black Box



Machine Learning Scope



*Deep Learning, 2016, by Ian Goodfellow, Yoshua Bengio, Aaron Courville

Game Theory to the Rescue

- Perspective of viewing any explanation of a model's prediction as model itself (explanation model)

- The model is based on Cooperative / Coalitional game theory concept
- -- Shapley Value

- A prediction is explained by assuming that each feature of the instance value is a player in the game where prediction is the payout

- Shapley value tells us how to fairly distribute the payout amongst the features

Shapley Value Estimation

Shapley value of a feature is its contribution to the payout, weighted and summed over all possible feature value combinations:

$$\phi_j(val) = \sum_{S \subseteq \{x_1, \ldots, x_p\} \setminus \{x_j\}} \frac{|S|!\,(p - |S| - 1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S) \right)$$

S : Subset of features used in the model
x : Vector of feature values of an instance to be explained
p: number of features

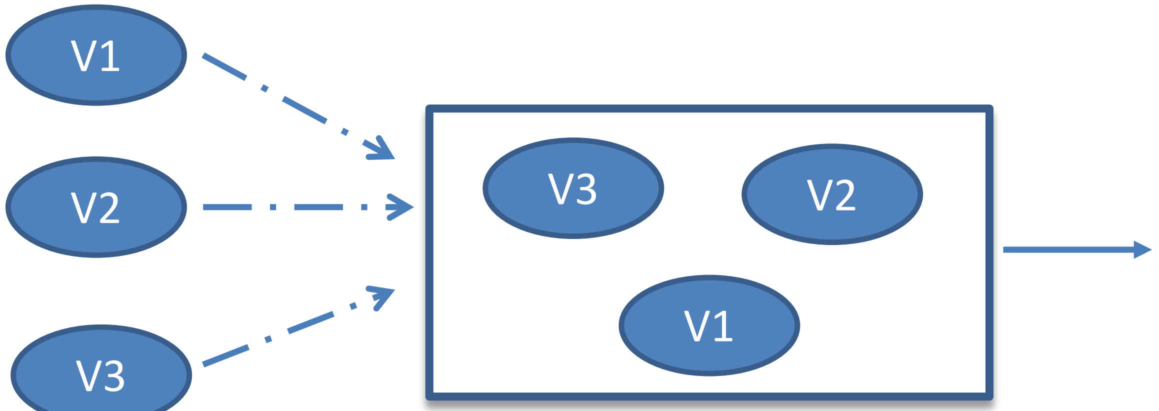
Simple Example

- Machine learning model contains 4 features, x1, x2, x3, x4
- Evaluate the prediction for the coalition comprising x1, x3

$$val_x(S) = val_x(\{x_1, x_3\}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}\,(x_1, X_2, x_3, X_4) d\mathbb{P}_{X_2 X_4} - E_X(\hat{f}\,(X))$$

- Resembles to feature contributions in linear methods

Shapely Value: Intuitive Explanation



Payout = Prediction

Room (Black Box Model)

Shapley Value: Average change in the prediction that the coalition already in the room receive when the feature value joins them

Shapely Value Estimation Algorithm

Evaluation of all possible coalitions are computationally expensive as the number of features grow

Monte Carlo sampling based approximation algorithm

- Output: Shapley value for the value of the j-th feature
- Required: Number of iterations M, instance of interest x, feature index j, data matrix X, and machine learning model f
- For all m = 1,…,M:
  - Draw random instance z from the data matrix X
  - Choose a random permutation o of the feature values
  - Order instance x: $x_o = (x_{(1)}, \ldots, x_{(j)}, \ldots, x_{(p)})$
  - Order instance z: $z_o = (z_{(1)}, \ldots, z_{(j)}, \ldots, z_{(p)})$
  - Construct two new instances
  - $x_{+j} = (x_{(1)}, \ldots, x_{(j-1)}, x_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
  - $x_{-j} = (x_{(1)}, \ldots, x_{(j-1)}, z_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
  - $\phi_j^m = \hat{f}\left(x_{+j}\right) - \hat{f}\left(x_{-j}\right)$
- Compute Shapley value as the average: $\phi_j(x) = \frac{1}{M} \sum_{m=1}^{M} \phi_j^m$

Case Study

- Data is related to direct marketing campaigns of a Portuguese Banking Institution (https://archive.ics.uci.edu/ml/datasets/bank+marketing)
- Data was also posted on Kaggle in an Analytics Challenge by Singapore Actuarial Society
- Problem is to model whether the client subscribed a Term Deposit or not (binary outcome) post the direct marketing campaign
- Technology used: SAS (Data preparation) and R (Model building and explanation)

Variables/Attributes in the Model

Total 17 attributes including

- **Client details** - Age, Job type etc.
- **Campaign details** - Communication type, last contact etc.)
- **Macroeconomic factors** - Consumer price index, Euribor 43 months rate etc.

Modelling Process

Data pre-processing:

- Variable transformation and missing value imputation
- Data was split as training (80%) and validation (remaining 20%)
- Classification algorithm used: Random Forest Classifier

Results and Discussion:

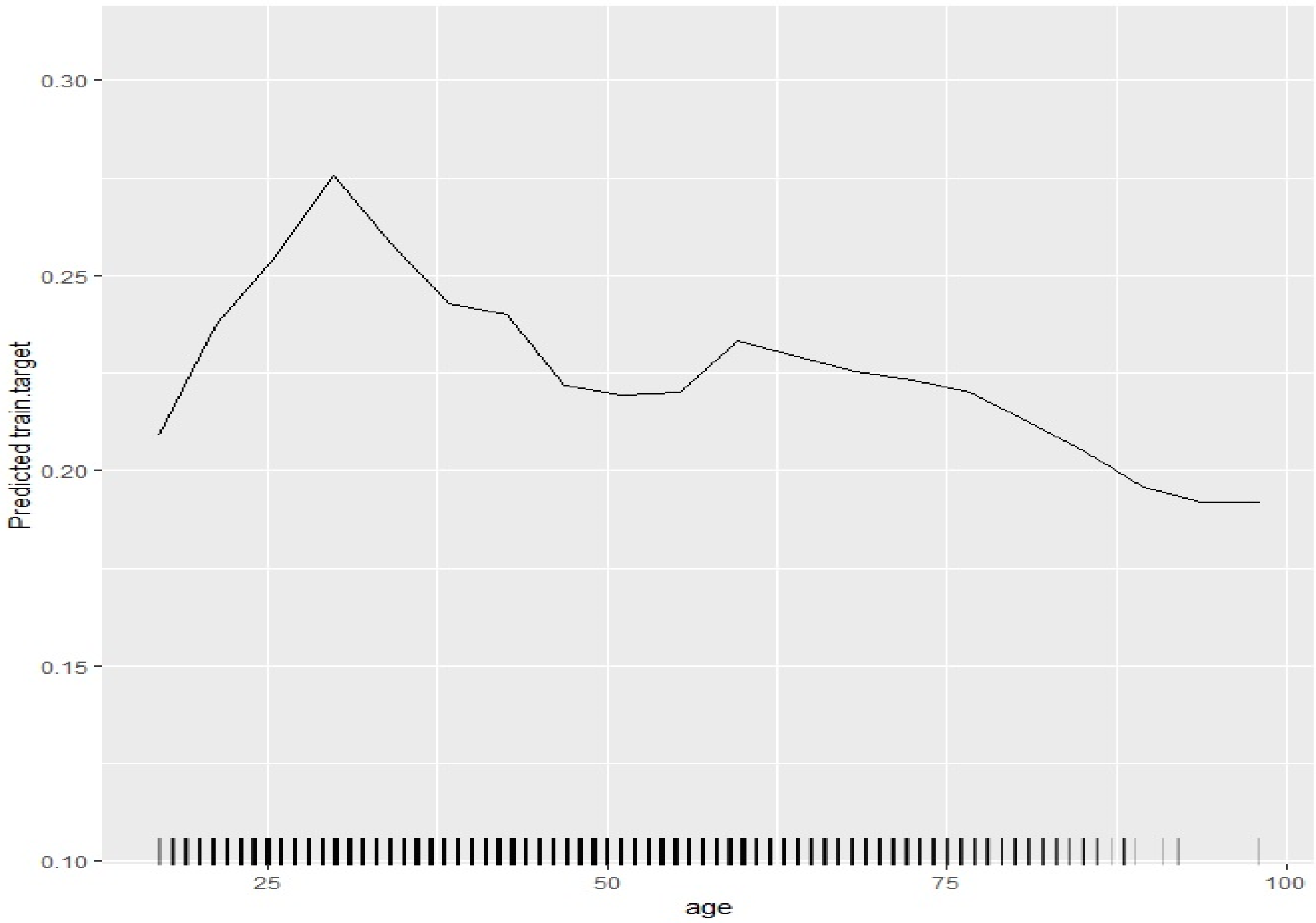
- We received consistent accuracy of @85% on both the splits

.

```
               Type of random forest: classification
                     Number of trees: 1200
No. of variables tried at each split: 15

        OOB estimate of  error rate: 13.81%
Confusion matrix:
       no  yes class.error
no  23227 2406  0.09386338
yes  1588 1705  0.48223504
                Test set error rate: 14.53%
Confusion matrix:
      no  yes class.error
no  8865 1003   0.1016417
yes  613  640   0.4892259
```

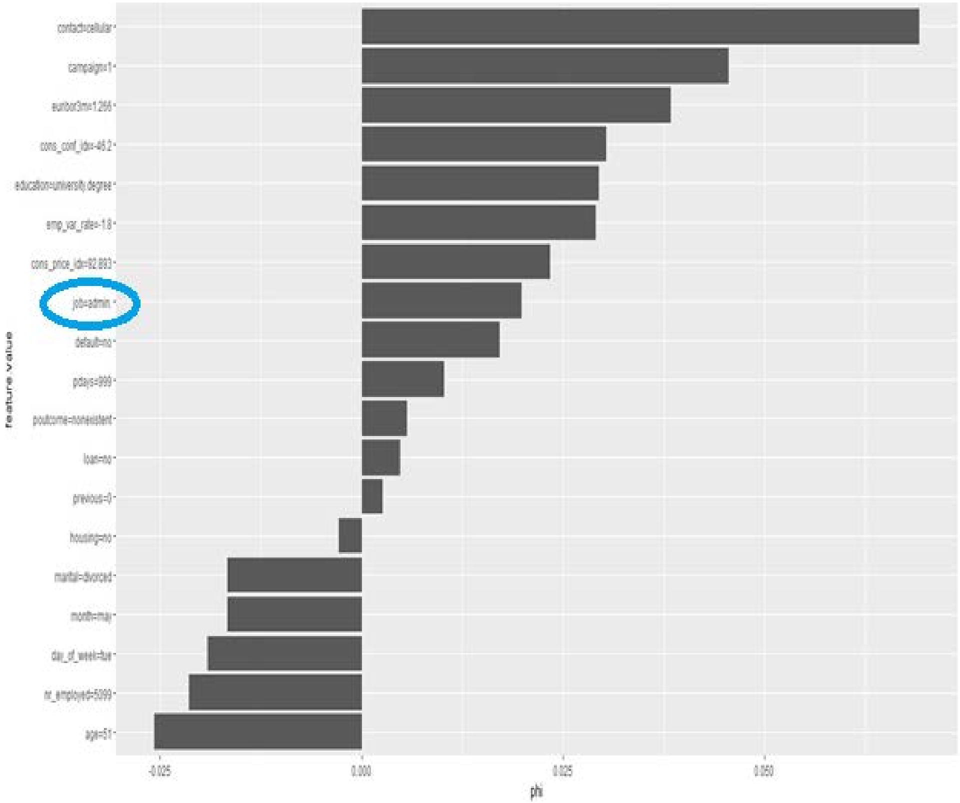
Variable Effect - Age



Model Explanation for One Client

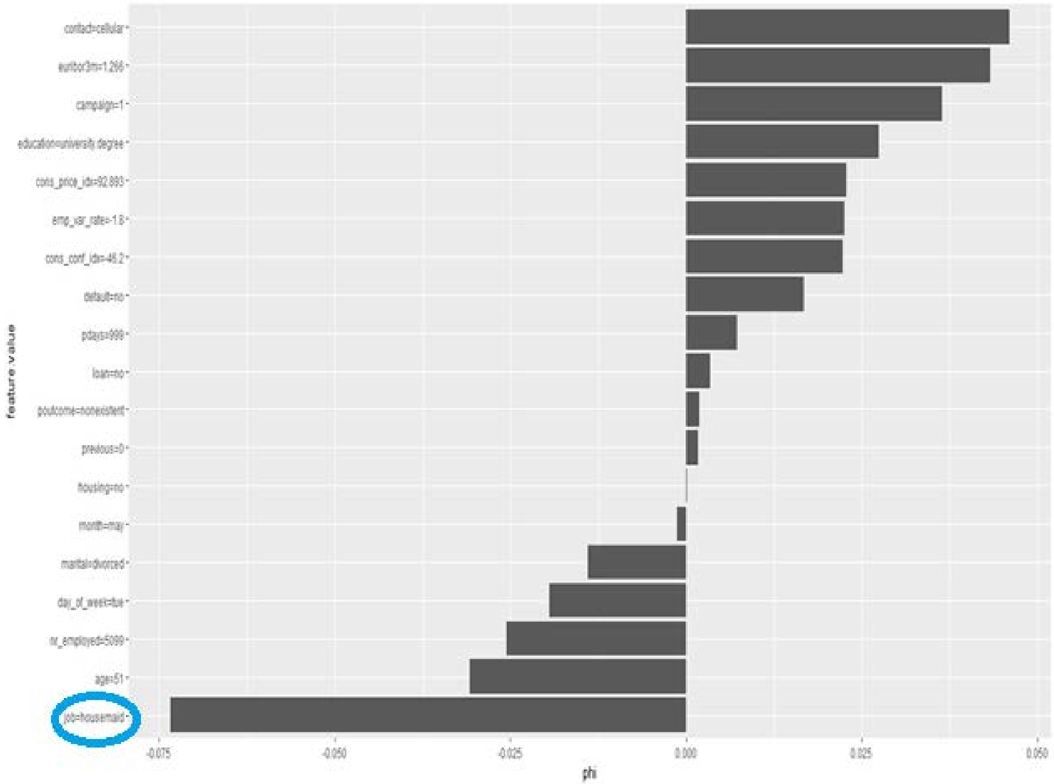
Job: Admin

Actual Prediction: 0.44 **Avg. Prediction: 0.25**



Job: Housemaid

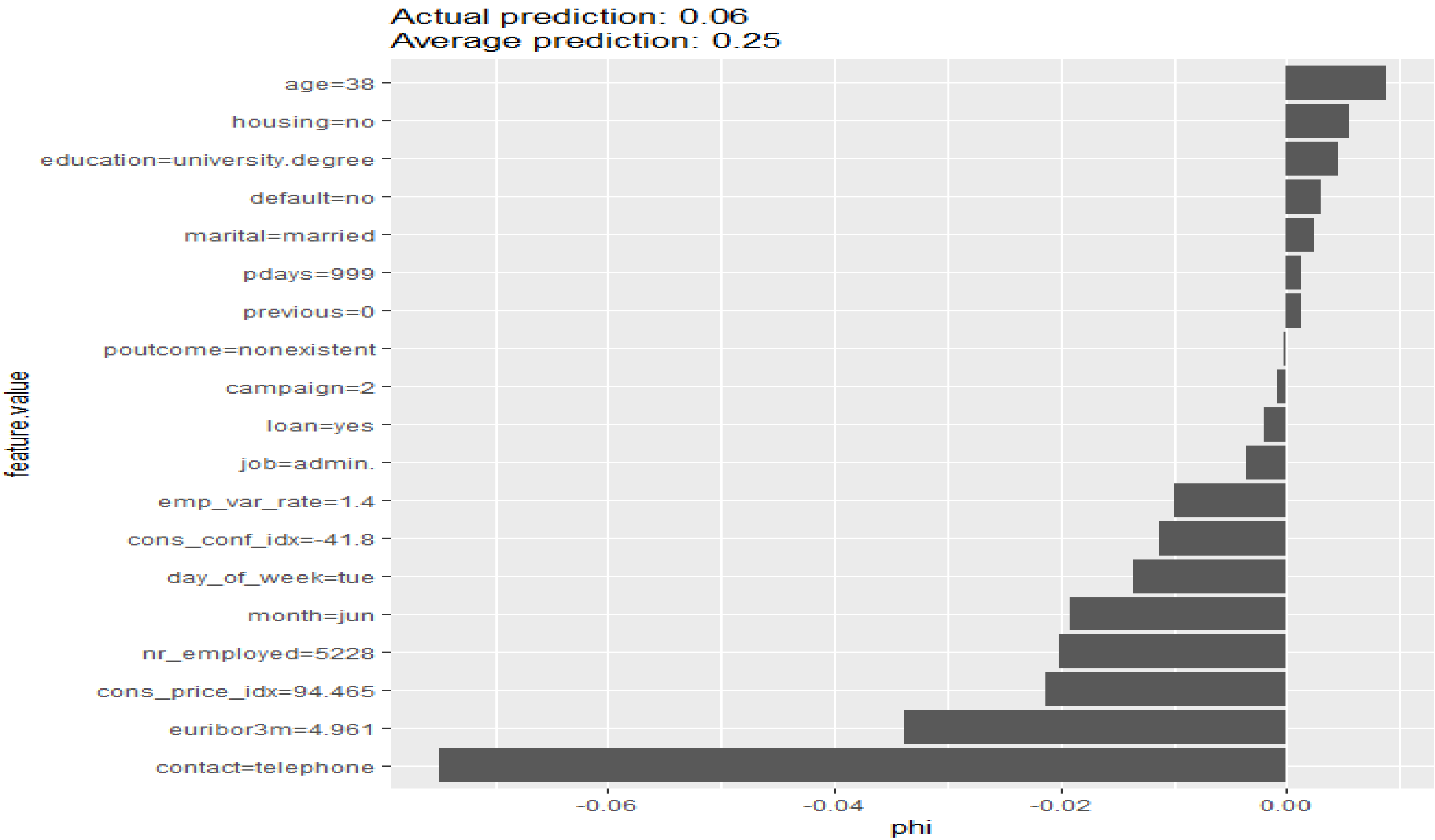
Actual Prediction: 0.3 **Avg. Prediction: 0.25**



20th Global Conference of Actuaries

4th - 6th March, 2019 | Mumbai, India

Model Explanation for One More Client



20th Global Conference of Actuaries
4th - 6th March, 2019 | Mumbai, India

Alternative Model Explanation Methods and Our Perspective

Alternative Explanation Methods

- Local Interpretable Model-agnostic Explanations (LIME)
- -- kind of surrogate model
- DeepLIFT
  - -- Specific to deep learning models
- Layerwise Relevance Propagation
- -- Specific to deep learning models
- Example based approaches

  -- Counterfactuals, Adversarial examples

Viewpoint and Way forward

- Of all the methods we studied, Shapley value based approach has sound theoretical basis and can augment the human decision making process nicely
- It is computationally expensive as the number of features grow
- Model Explainability is going to get more important
- Regulations (e.g. GDPR) are not supporting automated decision making without any human intervention
- Important to have analysis backed by explanation power

Institute of
Actuaries of India
Statutory body under an Act of Parliament

Expanding the horizon, strengthening the core

20th Global Conference of Actuaries

4th - 6th March, 2019 | Mumbai, India

THANK YOU

March 06, 2019