3rd Place Solution

Decision-Tree Aggregated Features and Hybrid Bin-Classifier/Quantile-Regressor

Yota Hashizume (hzume)

Notation

- $x \in \mathbb{R}^n$: Event features (e.g., `PRI_had_pt`, `DER_pt_tot`).
- $y \in \{0,1\}$: Event label (1 = signal, 0 = background signal).
- $\nu \in \mathbb{R}^6$: Nuisance parameters
- $\mu \in \mathbb{R}$: Signal strength (Estimation target)
- $\{(x_{ij}, y_{ij})\}_{j=1}^{M_i} \sim P(\nu_i, \mu_i)$: Samples under nuisance parameters ν_i and μ_i.

Overview

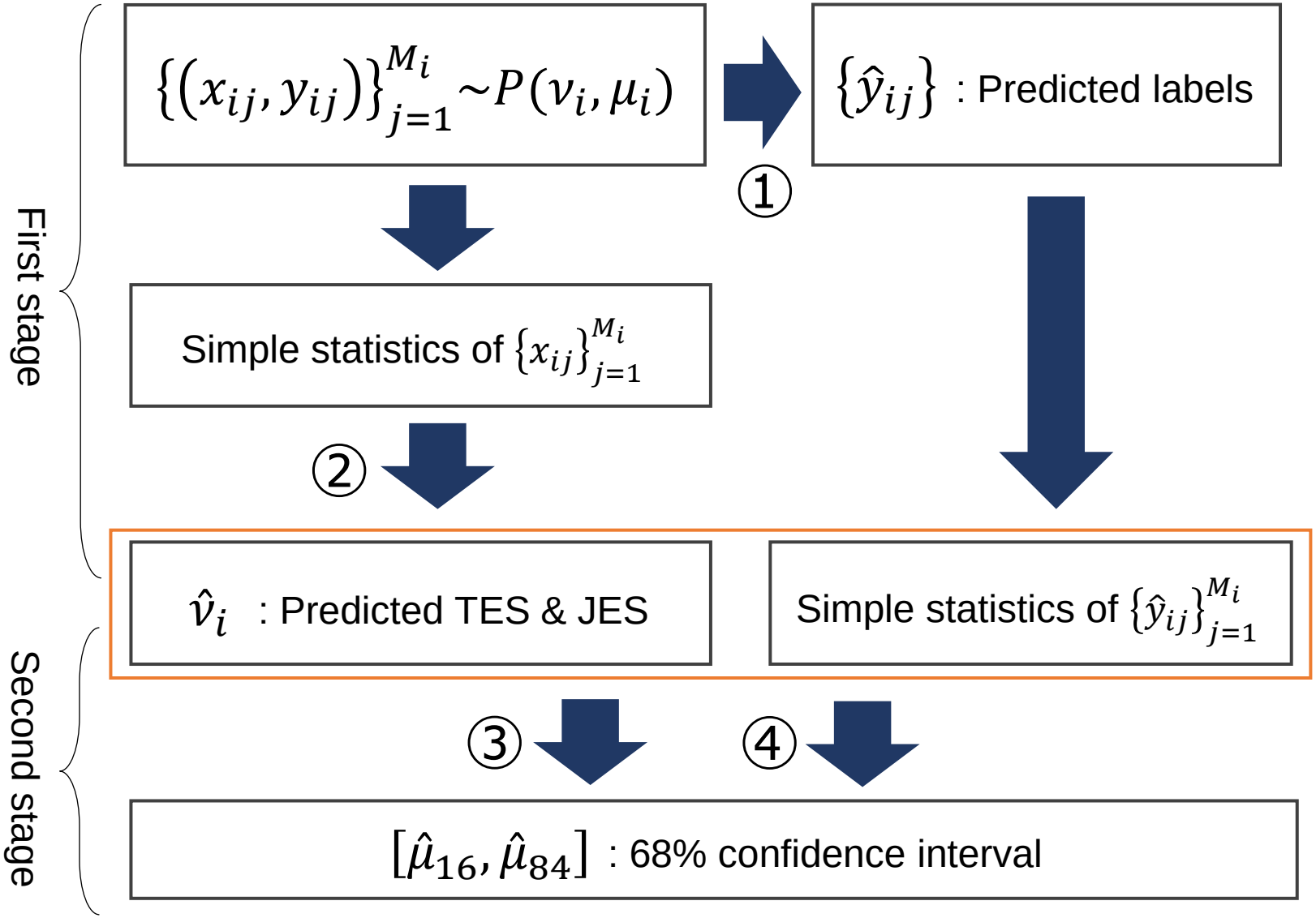
- 2-stage, GBDT-based model
- GPU-free

<u>First stage</u>

- Aggregated features
- Used 2 models (①,②)

<u>Second stage</u>

- Estimate 68% confidence interval by using aggregated features
- Used 2 models (③,④) and merged their outputs



First stage: Label estimator (Model ①)

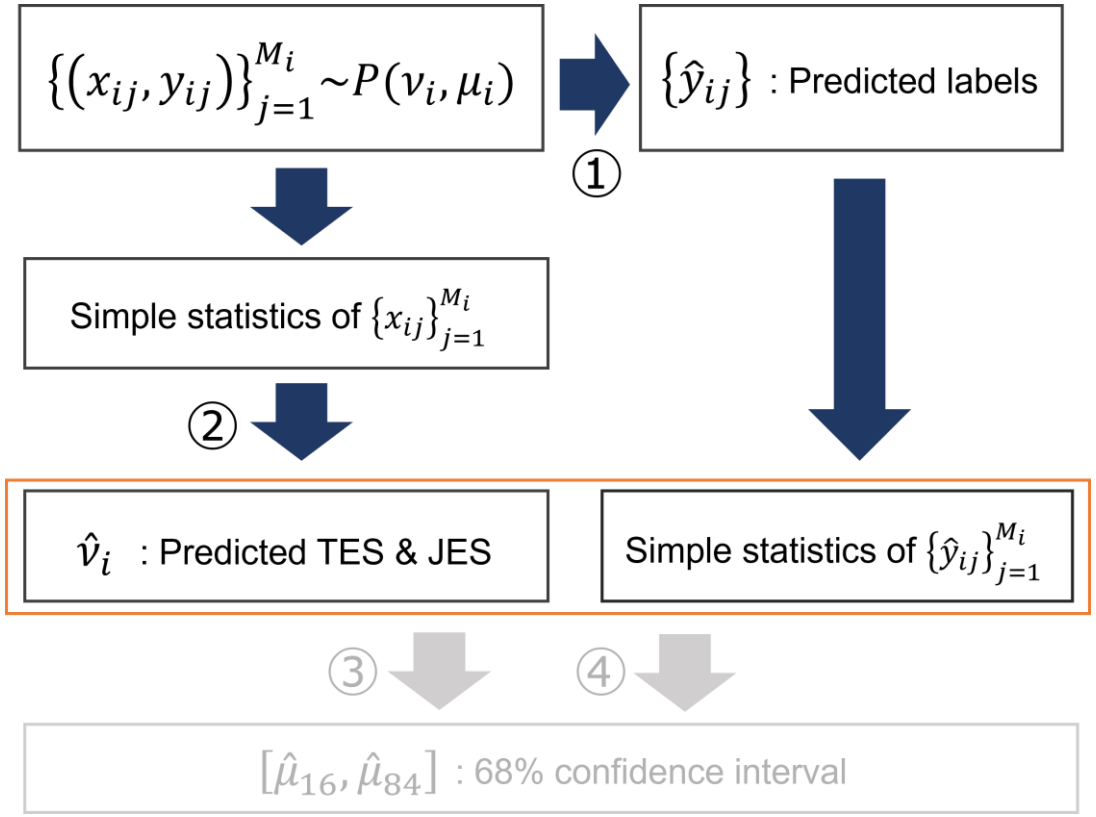
① : Label estimator

Training phase

- Trained by events under random nuisance parameters to predict their label
- $\hat{y} = f(x),\ (x, y) \sim \int p(x, y|\nu, \mu) p(\nu) d\nu$

Estimation phase

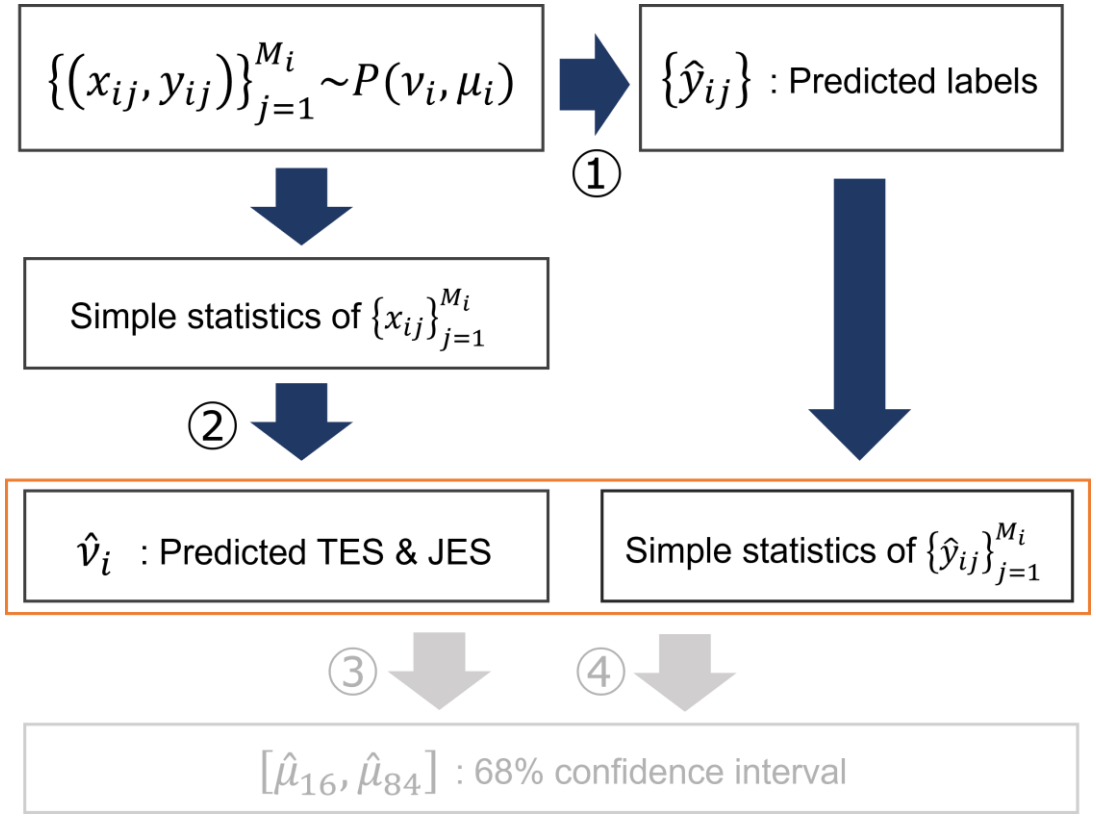
- $\{(x_j, y_j)\}_{j=1}^{M} \sim P(\nu, \mu)$
- Statics of predicted labels $\{\hat{y}_j\}_{j=1}^{M}$
  - Mean, variance, kurtosis, skewness
  - $\frac{i}{256}$ - quantile for $i = 0, \ldots, 256$
    (important to describe shape of the distribution)



First stage: TES & JES estimator (Model ②)

② : TES & JES estimator

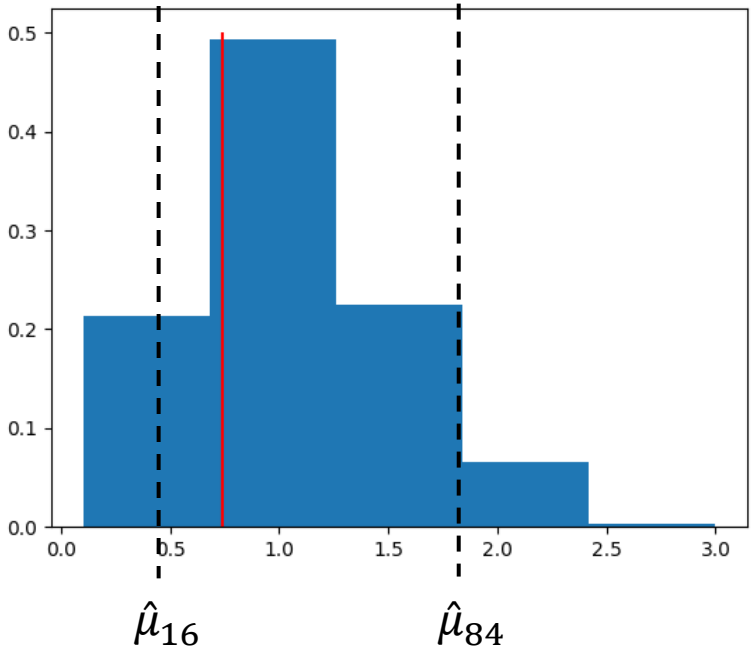
- TauEnergyScale (TES), JetEnergyScale (JES)
  - The factors applies to some features
- Statics of raw features $\{x_{ij}\}_{j=1}^{M_i}$
  - Mean, variance, kurtosis, skewness
- Predict TES and JES from statistics of features
- (Soft_MET was relatively difficult to predict in this way)



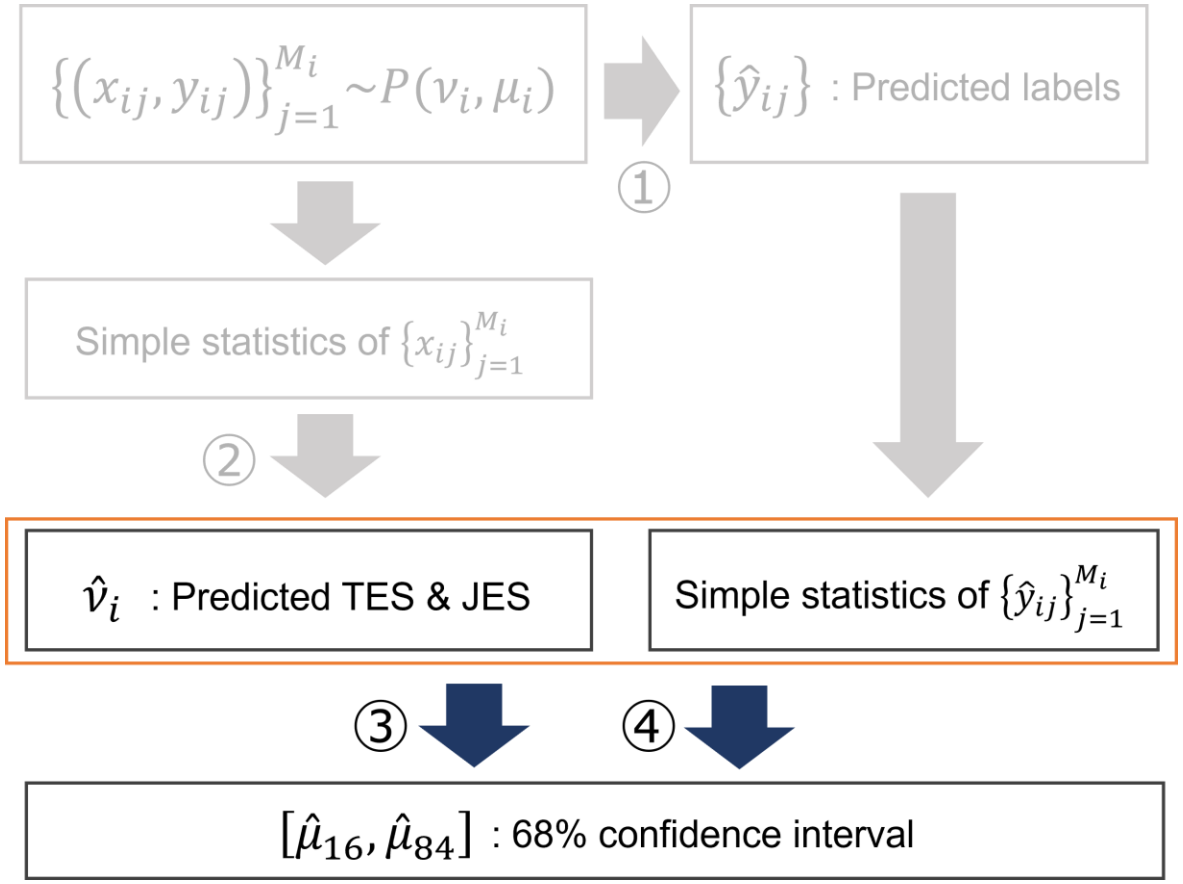
Second stage: Bin-Classifier (Model ③)

③ : Binned regression classifier

- Divide μ's range into 5 bins and predict its bin
- Treat the output as a distribution, then find the smallest interval that covers 68% of it.



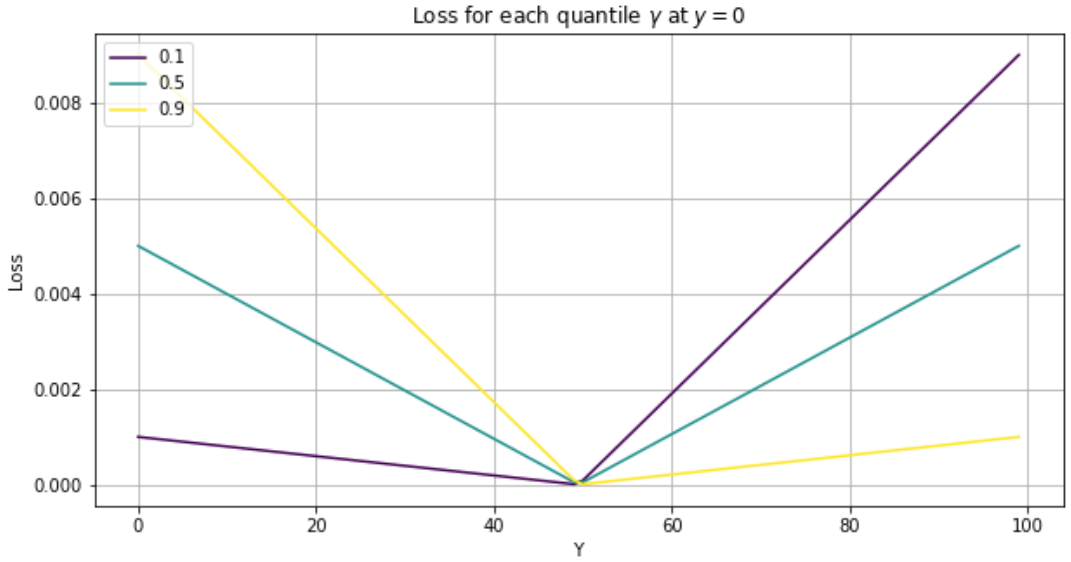
Example output of bin classifier

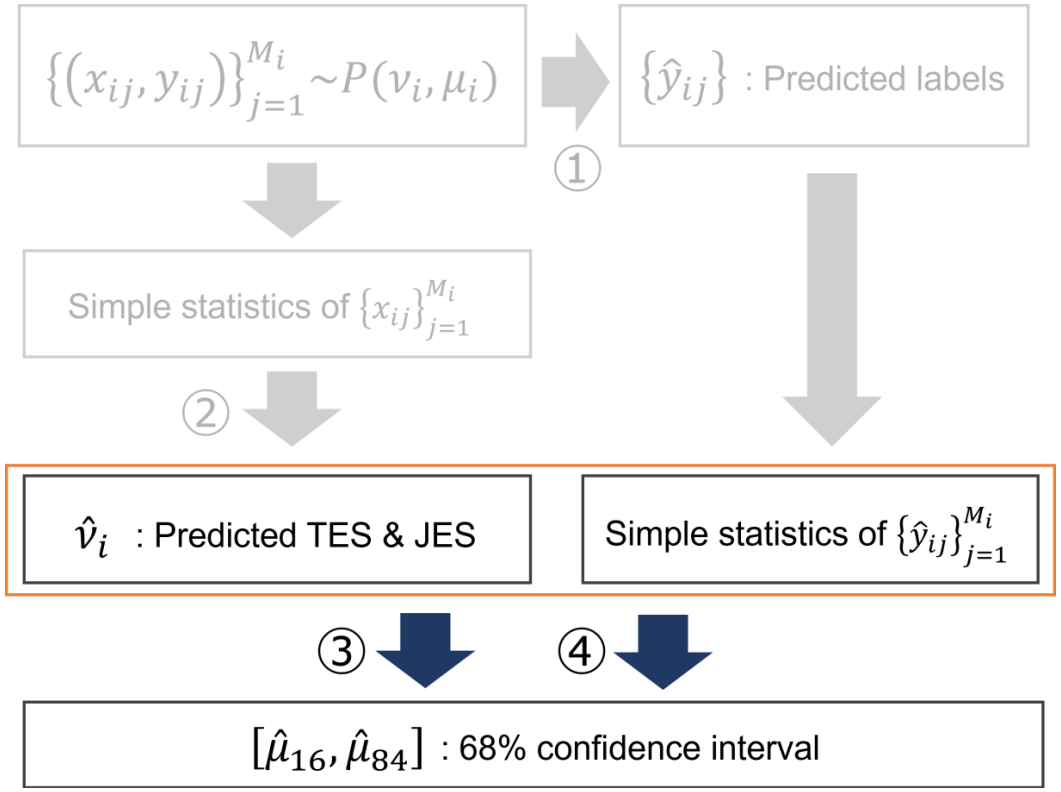


Second stage: Quantile-Regressor (Model ④)

④ : Quantile regressor

- Pinball loss function enables us to predict confidence interval directly
- You can use it easily by using LightGBM

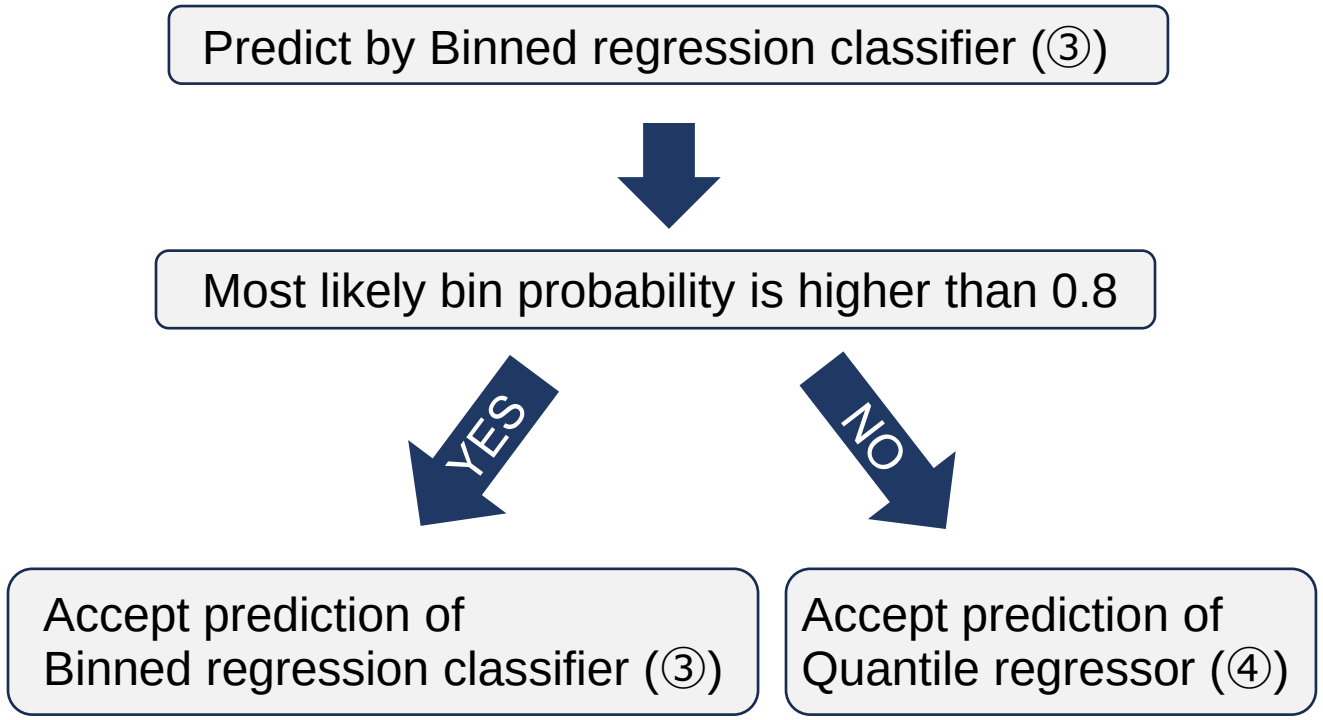


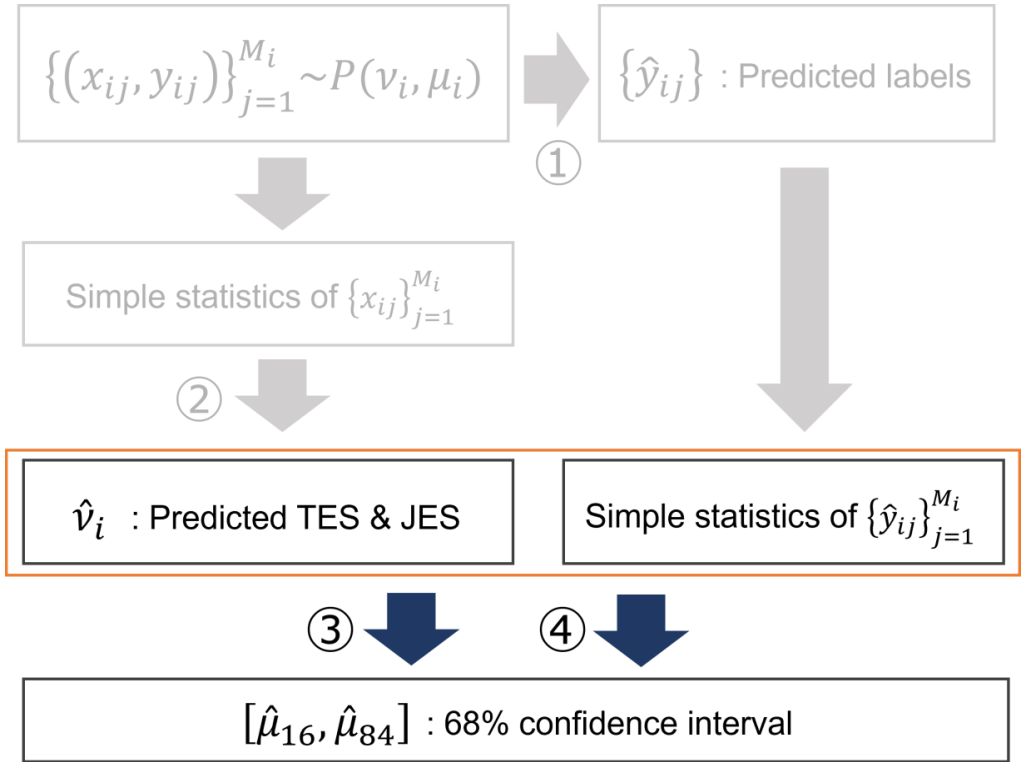


Second stage; Merging ③ and ④

Observation:

Binned regression classifier (③) can make better interval when its confidence is high.





Next steps

- Try a DNN approach
  - Deep sets, Set transformer, …
- Use bigger dataset
  - I used small dataset in this solution, but there is bigger (x1000~) dataset.
  - There will be computational cost issues, so need to make some adjustments.

Thanks to the host and the audience!