Capturing the power of ensemble learning using GLM and Artificial Neural Network for insurance pricing

S R Pranav Sai
Doctoral Research Scholar, SSSIHL

srpranavsai@sssihl.edu.in

+91 - 7358287929

GLM vs ANN

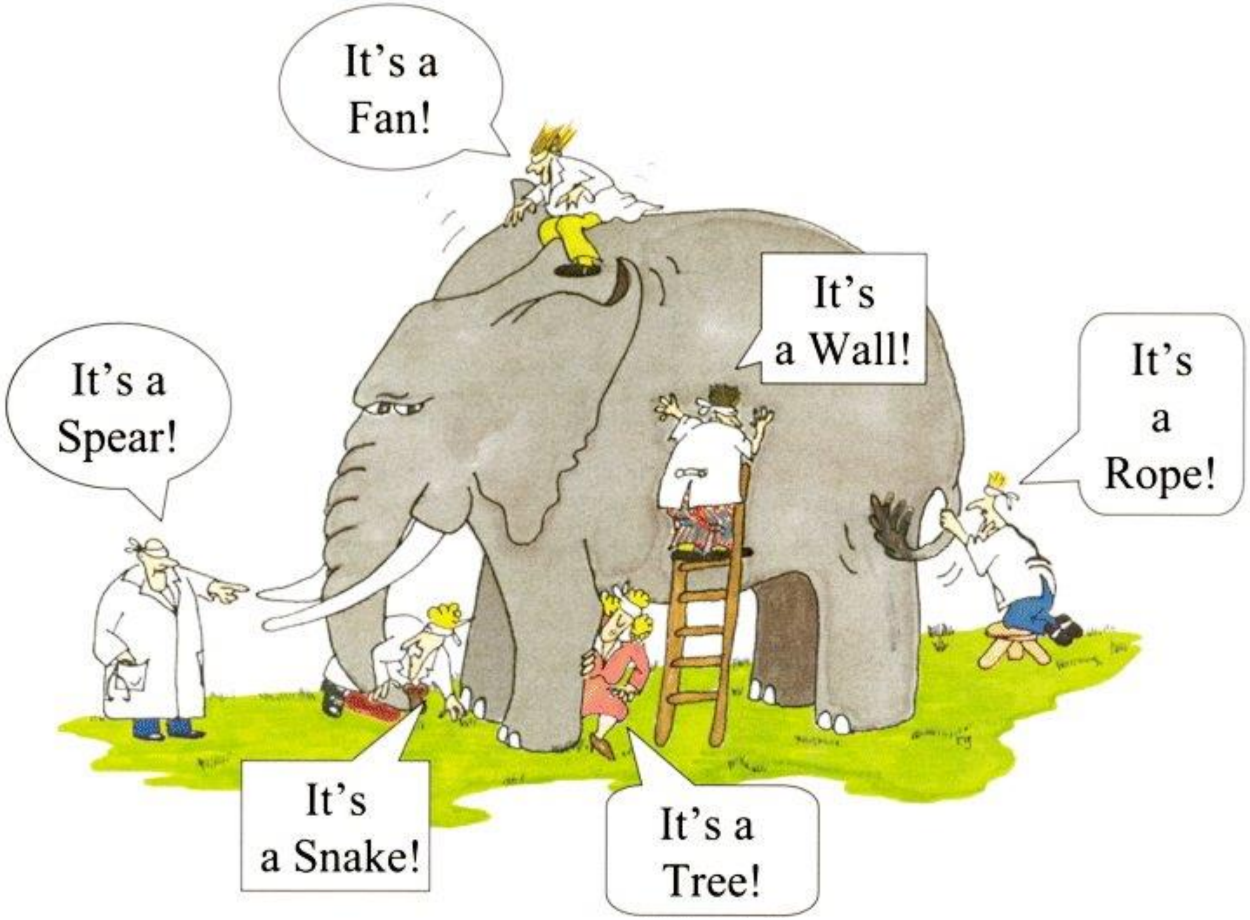
GLM

- Extension of the classical linear models
- Transparency in operations
- Ease of understanding
- Easy to explain
- Lesser time taken to fit a model
- Widely used in Non-Life Insurance pricing

ANN

- Inspired from the functioning of a human brain
- Effective pattern recognition ability
- Automatic variable interaction
- Flexibility in choosing weightage of the variables
- Lesser time to define the model topology
- More flexible

Ensemble Learning



The best model "knows less" about the data, than all others "weak models" combined.

https://www.patheos.com/blogs/driventoabstraction/2018/07/blind-men-elephant-folklore-knowledge/

The Data

	IDpol	ClaimNb	Exposure	Area	VehPower	VehAge	DrivAge	BonusMalus	VehBrand	VehGas	Density	Region
0	1.0	1	0.10	D	5	0	55	50	B12	Regular	1217	R82
1	3.0	1	0.77	D	5	0	55	50	B12	Regular	1217	R82
2	5.0	1	0.75	B	6	2	52	50	B12	Diesel	54	R22
3	10.0	1	0.09	B	7	0	46	50	B12	Diesel	76	R72
4	11.0	1	0.84	B	7	0	46	50	B12	Diesel	76	R72

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 678013 entries, 0 to 678012
Data columns (total 12 columns):
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   IDpol       678013 non-null  float64
 1   ClaimNb     678013 non-null  int64
 2   Exposure    678013 non-null  float64
 3   Area        678013 non-null  object
 4   VehPower    678013 non-null  int64
 5   VehAge      678013 non-null  int64
 6   DrivAge     678013 non-null  int64
 7   BonusMalus  678013 non-null  int64
 8   VehBrand    678013 non-null  object
 9   VehGas      678013 non-null  object
 10  Density     678013 non-null  int64
 11  Region      678013 non-null  object
dtypes: float64(2), int64(6), object(4)
memory usage: 62.1+ MB
```

We work with the **French Motor third party** liability dataset freMTPL2freq.

The dataset has **678,013** rows representing individual claim records and **12** columns representing the different features of those records.

Naive Model - Intercept Only

```
# Intercept only model
# claim frequency
cf <- sum(learn$ClaimNb)/sum(learn$Exposure)
learn$fit.cf <- cf*learn$Exposure
test$fit.cf <- cf*test$Exposure
```

Poisson Deviance Learn – 33.09%
Poisson Deviance Test – 32.79%

The Benchmark 'GLM2' Model

```
d.glm2 <- glm(ClaimNb ~ VehPowerGLM + VehAgeGLM
              + BonusMalusGLM + VehBrand + VehGas
              + DensityGLM + Region + AreaGLM
              + DrivAge + log(DrivAge) +  I(DrivAge^2)
              + I(DrivAge^3) + I(DrivAge^4),
                data=learn, offset=log(Exposure),
              family=poisson())
```

Poisson Deviance Learn – 31.23%
Poisson Deviance Test – 31.14%

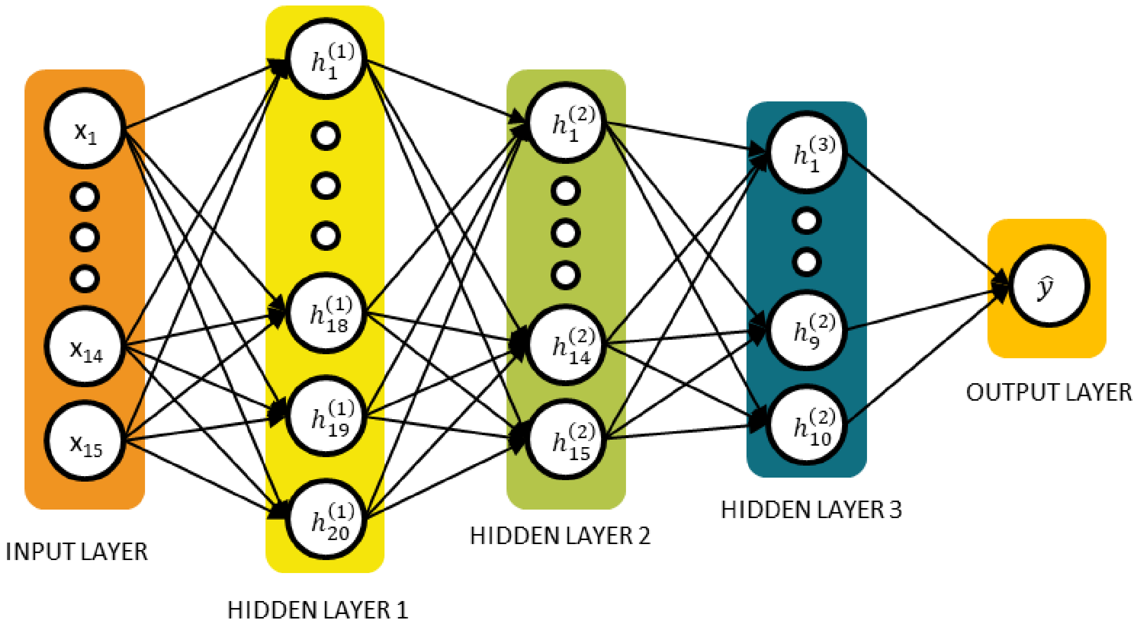
Schelldorfer, Jürg and Wuthrich, Mario V., Nesting Classical Actuarial Models into Neural Networks (January 22, 2019).

Evaluation of Models

```
# Function Benchmark.GLM2: Improvement in Poisson Deviance on test set compared to GLM2-INT-
Improvement
Benchmark.GLM2 <- function(txt, pred, obs) {
  index <- ((PD(pred, obs) - PD(test$fit.cf, test$ClaimNb)) / (PD(test$fitGLM2, test$Claim
Nb) - PD(test$fit.cf, test$ClaimNb))) * 100
  sprintf("GLM2-Improvement-Index (PD test) of %s: %.1f%%", txt, index)
}
```

$$\boldsymbol{GLM2\ Improvement\ Index\ of\ a\ \ Given\ Model} = \frac{PD(Given\ Model) - PD(Intercept - Only\ Model)}{PD(GLM2) - PD(Intercept - Only\ Model)}$$

The Neural Network Model



Poisson deviance learn – 30.39%
Poisson deviance test – 30.30%
The GLM2 improvement index of the ANN model is 150.9%.

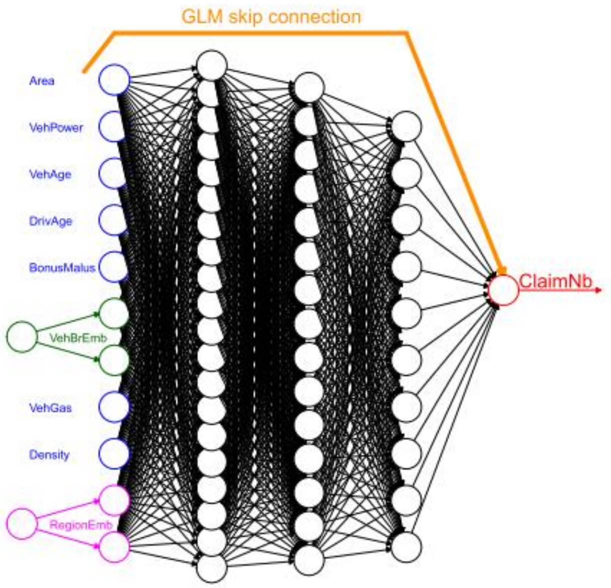
Combined Actuarial Neural Network - CANN

Nesting Classical Actuarial Models into Neural Networks

YES, WE CANN! (Combined Actuarial Neural Net)

Aim is to increase the acceptance of neural net modeling in the actuarial community

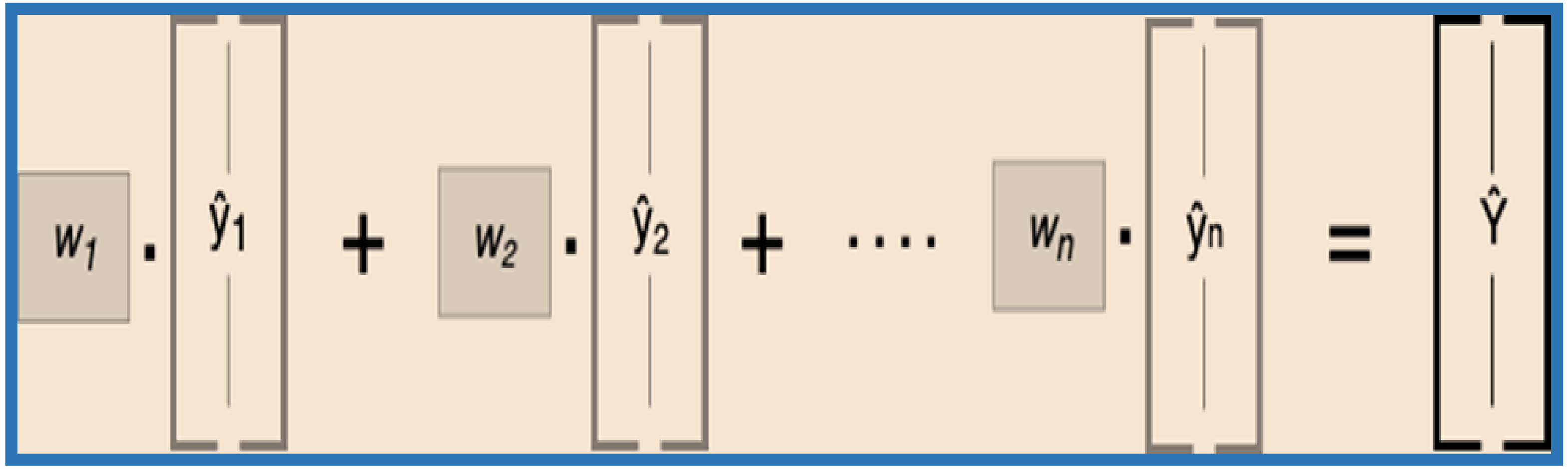
Neural nets may substantially improve classical actuarial models, if appropriately applied



Poisson deviance Learn – 30.58%
Poisson deviance Test – 30.22%
The GLM2 improvement index of the CANN model is 155.8%.

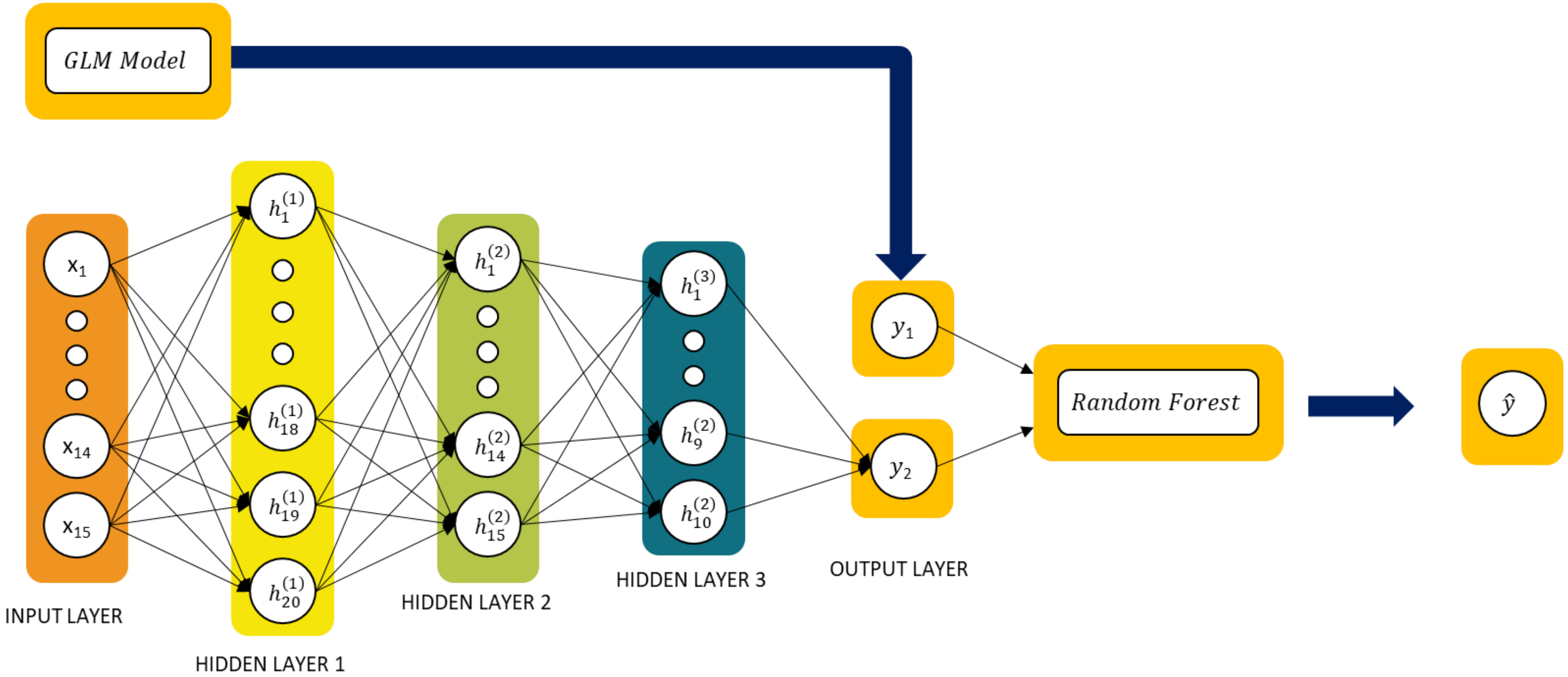
Schelldorfer, Jürg and Wuthrich, Mario V., Nesting Classical Actuarial Models into Neural Networks (January 22, 2019).

The Ensm1 Model – Weighted Average Ensemble



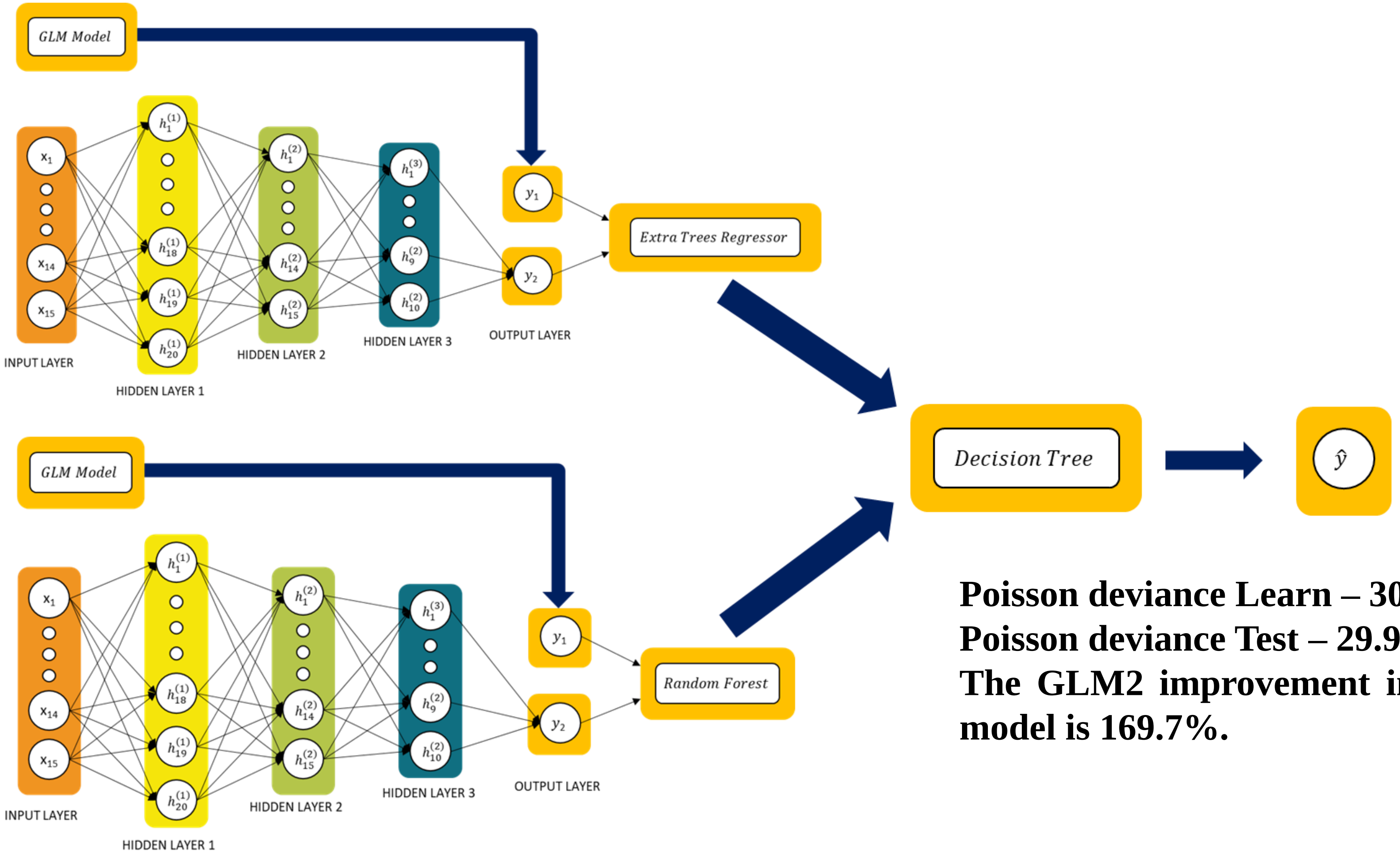
Poisson deviance Learn – 30.58%
Poisson deviance Test – 30.28%
The GLM2 improvement index of the ENSM1 model is 152.1%.

The Ensm2 Model – Stacked Ensemble



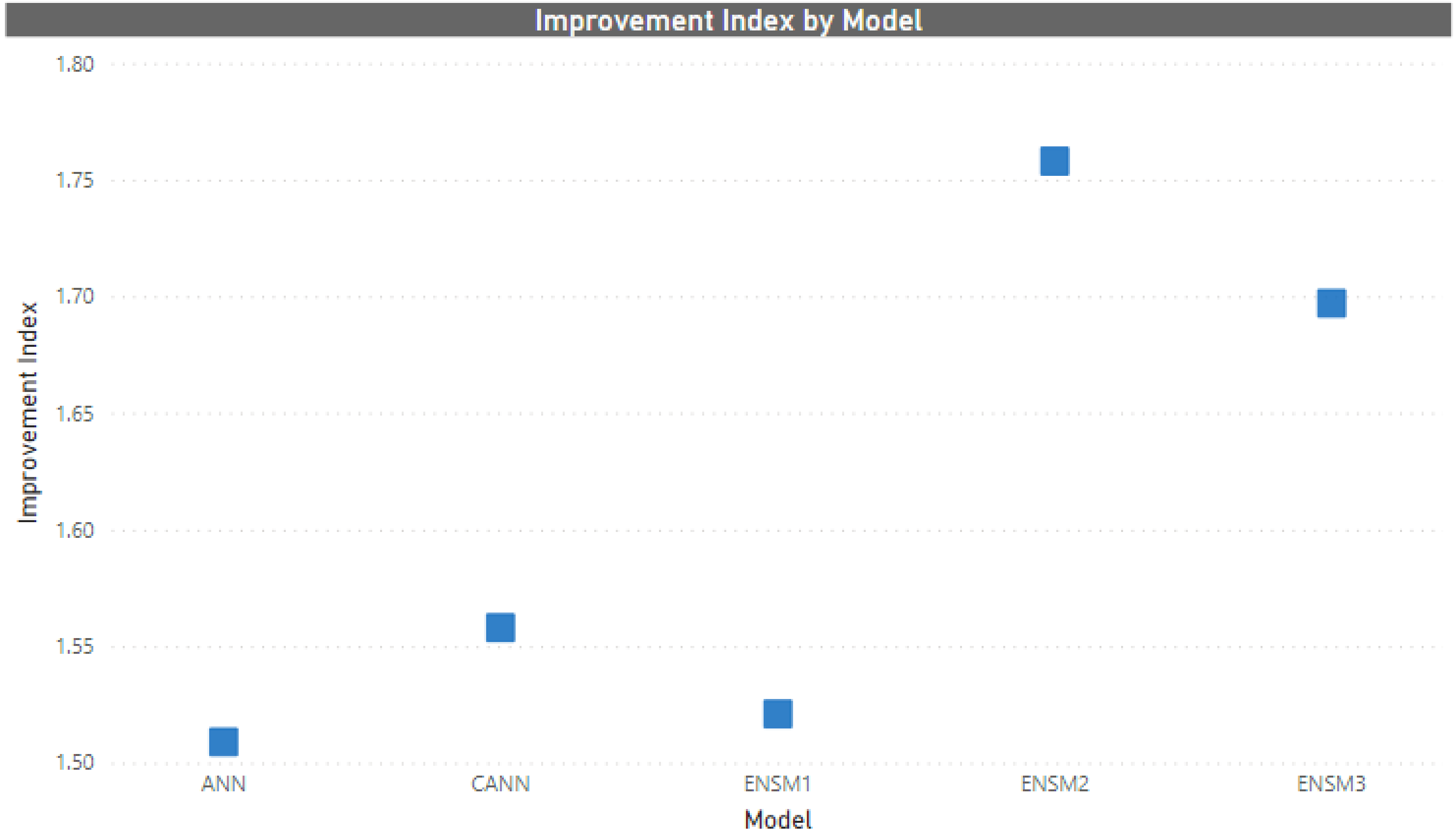
Poisson deviance Learn – 28.39%
Poisson deviance Test – 29.89%
The GLM2 improvement index of the ENSM2 model is 175.8%.

The Ensm3 Model - 3rd Level Stacked Ensemble



Poisson deviance Learn – 30.50%
Poisson deviance Test – 29.99%
The GLM2 improvement index of the ENSM3 model is 169.7%.

The Results



Improvement Index by Model
1.80
1.75
1.70
1.65
1.60
1.55
1.50
Improvement Index
ANN
CANN
ENSM1
ENSM2
ENSM3
Model

Acknowledgements

- Bhagawan Sri Sathya Sai Baba, Founder chancellor, SSSIHL
- Research Supervisors:
  - Satya Sai Mudigonda, Honorary professor (Actuarial Science), Department of Mathematics and Computer Science, SSSIHL
  - Dr. Pallav Kumar Baruah, Department of Mathematics and Computer Science, SSSIHL
  - Phani Krishna Kandala, Visiting faculty, SSSIHL

SRI SATHYA SAI INSTITUTE OF HIGHER LEARNING

PRASANTHI NILAYAM

सत्यं वद धर्मं चर