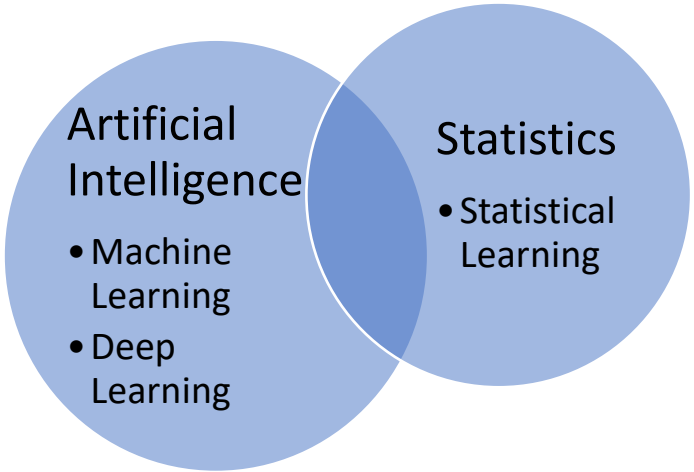
Artificial Intelligence
General Overview with Examples in Machine Learning & Deep Learning

Presented by:
Eric Munger, Ph.D.
Polytrauma System of Care, Health Science Researcher
VA Palo Alto Health Care System, Department of Veterans Affairs

Presented to:
Traumatic Brain Injury Model Systems, Analytic Special Interest Group

Statistics & Artificial Intelligence



- Objective:
  - What is AI?
  - What is Machine Learning?
  - What is Deep Learning (Neural Networks)?
  - Machine Learning vs Statistical Learning
  - Basic Concepts in Machine Learning
  - Example: Application of Random Forest Machine Learning
  - Example: Deep Learning & Neural Networks

Machine Learning (ML) & Deep Learning?

- Machine learning is a branch of artificial intelligence that enables algorithms to discover hidden patterns within datasets.
  - Typically called “Traditional Machine Learning” algorithms.
  - Maintain some interpretability -> Grey Box Modeling
  - Traditional machine learning requires data preprocessing.

- Deep learning is a type of artificial intelligence (AI) that uses artificial neural networks to learn how to process data and make decisions.
  - Inspired by the human brain.
  - Uses multiple layers of neural networks.
  - Doesn’t require significant data preprocessing.
  - Very difficult to interpret -> Black Box Modeling

Machine Learning & Statistical Learning

Statistical Learning vs Machine Learning		
	Statistical Learning	**Machine Learning**
Focus	Hypothesis testing	Predictive performance
Driver	Math, theories, hypothesis	Fitting parameters
Data size	Small, within reason	Big data
Data type	Structured	Structured, unstructured, semi-structured
Dimensions	Low dimensional	Low & High dimensional
Model selection	Parameter significance	Validation of predictive performance on test data
Inturpretability	High	Low
Strength	Define causal relationships	Prediction accuracy

Choosing the right AI algorithm

- Broadly speaking, **ML is under the umbrella of artificial intelligence** (AI)
  - **Deep Learning** is a **subset** of traditional **machine learning**

- Traditional machine learning - **significant effort** put into **data preprocessing**
  - **Normalizing, scaling, data cleaning, or feature engineering** before implementation

- In deep learning - effort is applied towards **tuning** and **optimizing the model after it's implemented**

Factors	Machine Learning	Deep Learning
Human Interpretation	Algorithms are logical	Likely impossible to interpret
Available Algorithms	Many	Few
Dataset	May be small or incomplete	Requires large datasets
Training Costs	Less time and computational costs	Requires significant computational effort
Feature Engineering	Must understand features that represent the data	Does not need to understand the features
Hardware	Easy to train on CPU	Requires CPU or GPU network
Accuracy	Lower accuracy	High accuracy
Feature Selection	Yes	No
Classification	Yes	Yes
Regression	Yes	Yes

Machine Learning Training Strategies

- The training strategy can be categorized based on the learning style the algorithms uses. Very much like statistical learning:

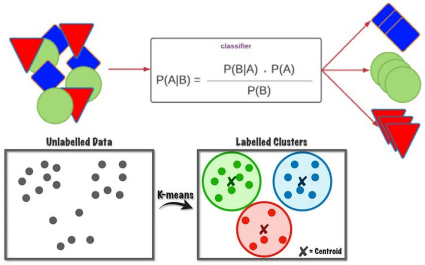
  - **Supervised Learning** Algorithms
    - Each record **includes** a classification or regression **label**
    - Ex: Y/N or time-variant quantity

  - **Unsupervised Learning** Algorithms
    - Accept **unlabeled data** - records do not have a known result
    - Identifies **structured patterns** within the input data
    - Allows for the **identification of unseen patterns**
    - Can be used to **organize data by the identified patterns**

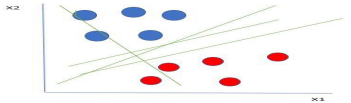
  - **Reinforcement learning**
    - Identify trends in data and make inferences without knowledge of correct answers.
    - A reward-and-punishment paradigm as they process data.
    - They learn from the feedback of each action and self-discover the best processing paths to achieve outcomes.

What are the most common and popular machine learning algorithms?

- Naïve Bayes Classifier Algorithm (Supervised Learning - Classification)
  - It allows us to predict a class/category, based on a given set of features, using probability.

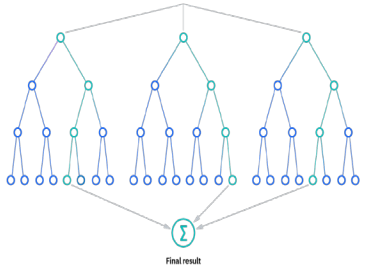


- K Means Clustering Algorithm (Unsupervised Learning - Clustering)
  - Used to categorize unlabeled data, i.e. data without defined categories or groups.
  - Finding groups within the data - iteratively assign each data point to one of K groups based on the features provided.



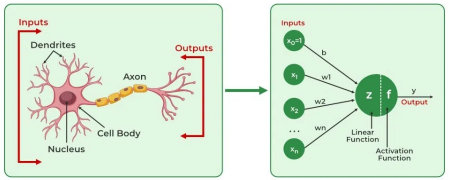
- Support Vector Machine Algorithm (Supervised Learning – Classification & Regression)
  - Separate data by hyperplanes - then build a model that assigns values based on the hyperplane.

- Decision Trees (Supervised Learning – Classification/Regression)
  - A tree structure that uses a branching method to illustrate every possible outcome of a decision.
  - Each node within the tree represents a test on a specific variable – each branch is the outcome of that test.



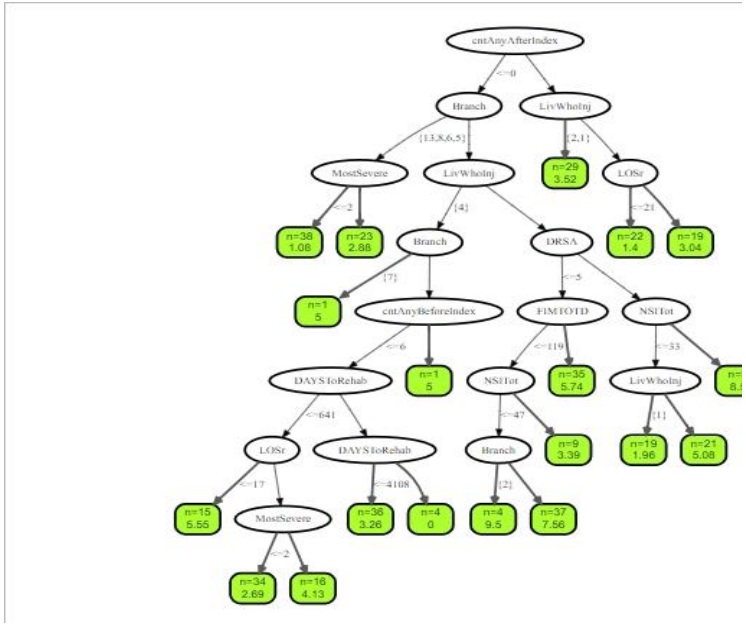
- Random Forests (Supervised Learning – Classification/Regression)
  - Ensemble learning method, combining multiple weak learners to generate better results.
  - Used as example 1: Resistant to overfitting, accuracy, wide data, feature importance, missing values

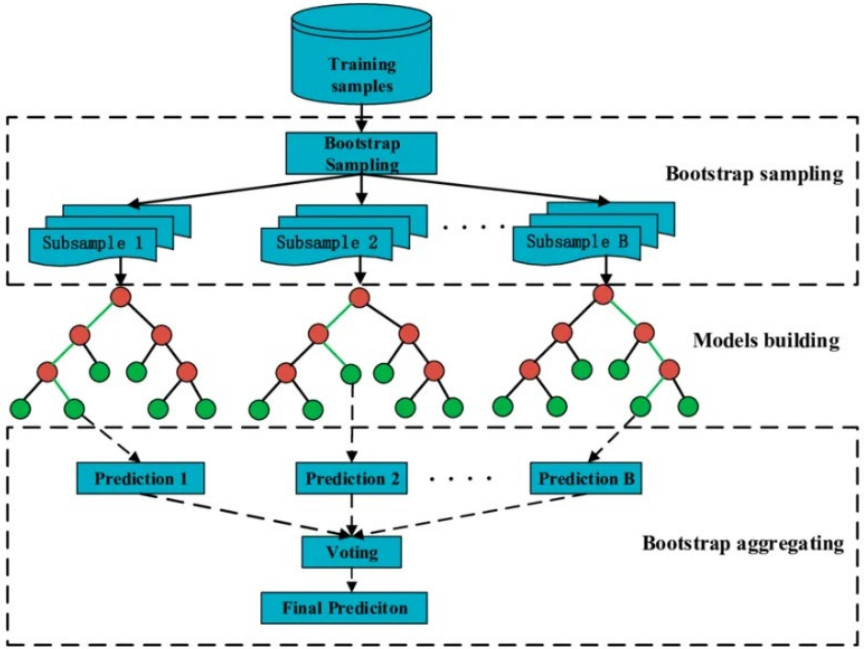
- Artificial Neural Networks (Reinforcement Learning)
  - ANNs - perceptrons arranged in a series of layers.
    - Inspired by the biological brain, and how they process information.
  - ANNs also learn by example and through experience
  - Used as example 2: modeling non-linear relationships, high-dimensional data or big data.



What is a Random Forest (RF) Machine Learning (ML) Model?

- A machine learning algorithm that combines the output of multiple decision trees (ensemble learning).
- Each tree is fit from a bootstrap sample (random selection with replacement) of the training observations.
- Out-of-bag (OOB) error is calculated by averaging the prediction error using only the observations that did not include that sample in the bootstrap sample.

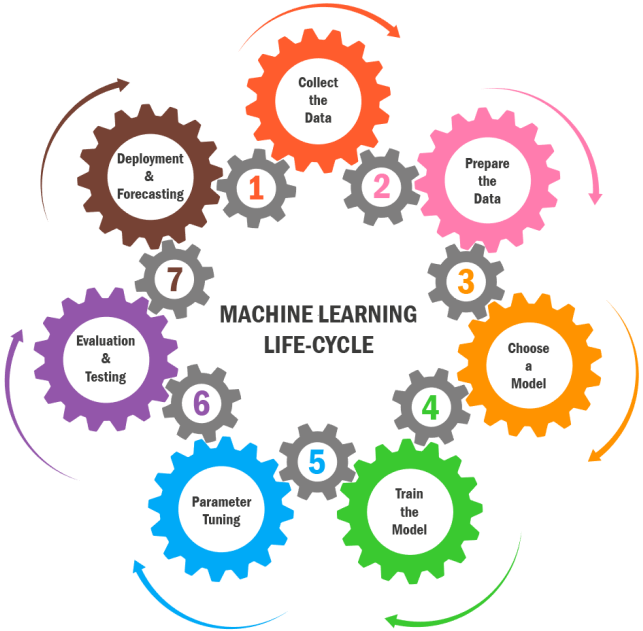




Example Decision Tree and RF workflow

Steps in the machine learning process

- Data collection
  - Capture data without processing and maintain an unaltered record of the original data.
- Data Preparation
  - Raw data is transformed into a format that is suitable for analysis.
  - Involves cleaning and validating the data.
  - Identifying any potential issues or errors.
- Model selection
  - A model is chosen.
- Model training
  - The prepared data is passed to the machine-learning model
  - Find patterns and make predictions.
- Model evaluation
  - This is a key step in the machine learning process. The evaluation process differs for supervised and unsupervised models.
- Parameter tuning
  - Different values of a parameter are tried and the value that optimizes.
  - Measured by predictive accuracy.
- Model Deployment
  - The model is deployed into production.
  - Used for real-world applications.



Example 1: Random forest machine learning

- Based on current VA TBIMS notification:
  - Machine Learning Identifies Characteristics of Retention in a Longitudinal Study of Veterans with Traumatic Brain Injury: A VA TBI Model System Study
- Machine Learning Identifies Characteristics of Retention in a Longitudinal Study
  - Aim 1: Using Form I and selected follow-up data (Lost & Date Last Followed), apply Random Survival Forest (RSF) to analyze time to event, i.e. loss of follow-up (LTFU). Identify important features related to LTFU.
  - For this work, LTFU is defined as a subject being lost from further participation in follow-up interviews and not returning.
  - Competing risks were not considered.

Traditional Statistical Survival Analysis & Key Concepts

Time to Event

The period from a defined starting point to the occurrence of the event of interest, such as death or disease recurrence.

- In our case, LTFU.

Censoring

A common feature in survival data where the event of interest has yet to occur for some subjects by the end of the study period but may occur later.

- LTFU as defined is right censored

Kaplan-Meier Estimator

A non-parametric method for estimating the survival function from censored data,

Provides a stepwise curve representing survival probability over time.

Cox Proportional Hazards Model

A widely used semi-parametric model that assesses the effect of covariates on survival time without assuming a specific baseline hazard function.

Random Survival Forest Machine Learning Algorithm

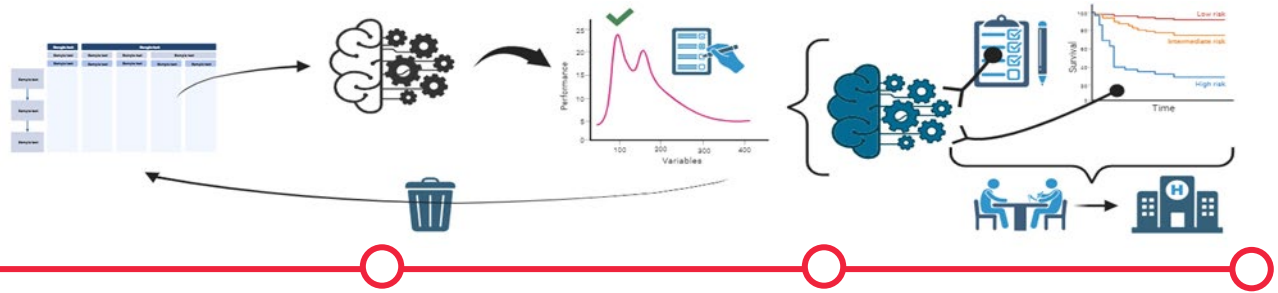
Key Components of RSF

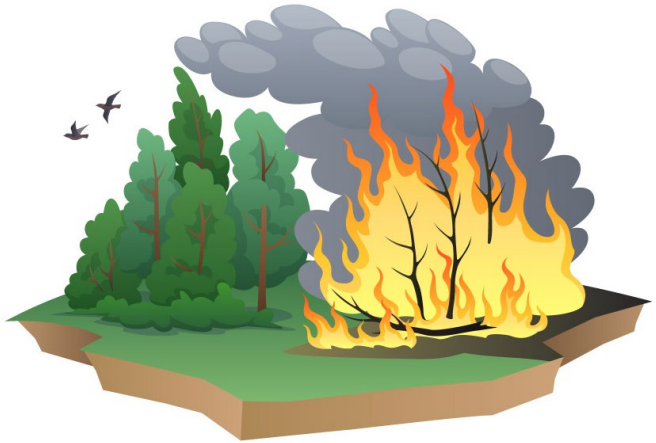
- RSF constructs decision trees during training and aggregates results to predict survival outcomes.
- Uses a survival-specific splitting criterion, the log-rank test, using the null hypothesis that there is no difference in survival between two groups.
- Employs out-of-bag samples to estimate prediction error, enhancing model performance assessment.
- Calculates variable importance scores to identify influential features for predicting survival outcomes.

Advantages Over Traditional Methods

- Non-parametric, avoiding assumptions about the distribution of survival times, allowing for complex modeling.
- Handles high-dimensional data effectively, performing implicit feature selection and resisting overfitting.
- Robust to right-censored data, maintaining accuracy in predictions without ad-hoc adjustments.
- Improved predictive accuracy due to ensemble nature, reducing variance and enhancing reliability.

Novelty: Using a random survival forest as a recursive feature selector to reduce dimensionality





- **Preprocess data**
- **Train and optimize a random survival forest model.**
  - **Rank and plot the OOB prediction error.**
  - **Save the model if OOB predictive error improves.**
  - **Eliminate the least important variable.**
- **Repeat the cycle until all prediction features are eliminated.**
- **Validate final model using test data**
- **predict survival probabilities, and error rates for test data.**
- **Analyze the results.**

Optimization Forest Fire:

- (Features = 197) X (Node Size = eles[10,15,20,25] = 4) X (Splits = floor(√197) = 14) X (Num Trees = 1001(1001+1)/2 = 501,501)
- ~ 9 hours on Dell Alienware (Core™ i9 CPU Processor, GeForce RTX 4090 (4080 GPU processors), and 64 GB RAM)
- **11,032 random forests consisting of ~5.5 X 10^9 decision trees grown and burned down.**

Dataset and Processing

Dataset Overview

The dataset includes 2,232 traumatic brain injury patients enrolled in the Traumatic Brain Injury Model Systems (TBIMS) longitudinal research project, which is the longest and largest longitudinal database on individuals with moderate -to-severe TBI.

Data Characteristics

The dataset included demographic information, clinical characteristics, and participant follow -up data over a specified observation period, with a follow -up time ranging from a minimum of 1 year to a maximum of 30 years.

Data Processing Steps

- 76 records where only administrative variables were collected were removed.
- 19 administrative variables were removed.
- 97 variables with little or no variance were deleted.
- Markers of data values unavailable were re-coded as missing.
- 177 variables were removed due to high missingness.
- Remaining missing values were imputed with mean or mode.
- Applied random undersampling to prevent overfitting.
- Split dataset into 80% for training (579X199) & 20% for testing (151X199).

RFE & Final Model

OOB prediction error optimized at ~1000 trees

Each model was initiated and hyperparameters were optimized to improve model performance:

- tree depth, and node splitting.

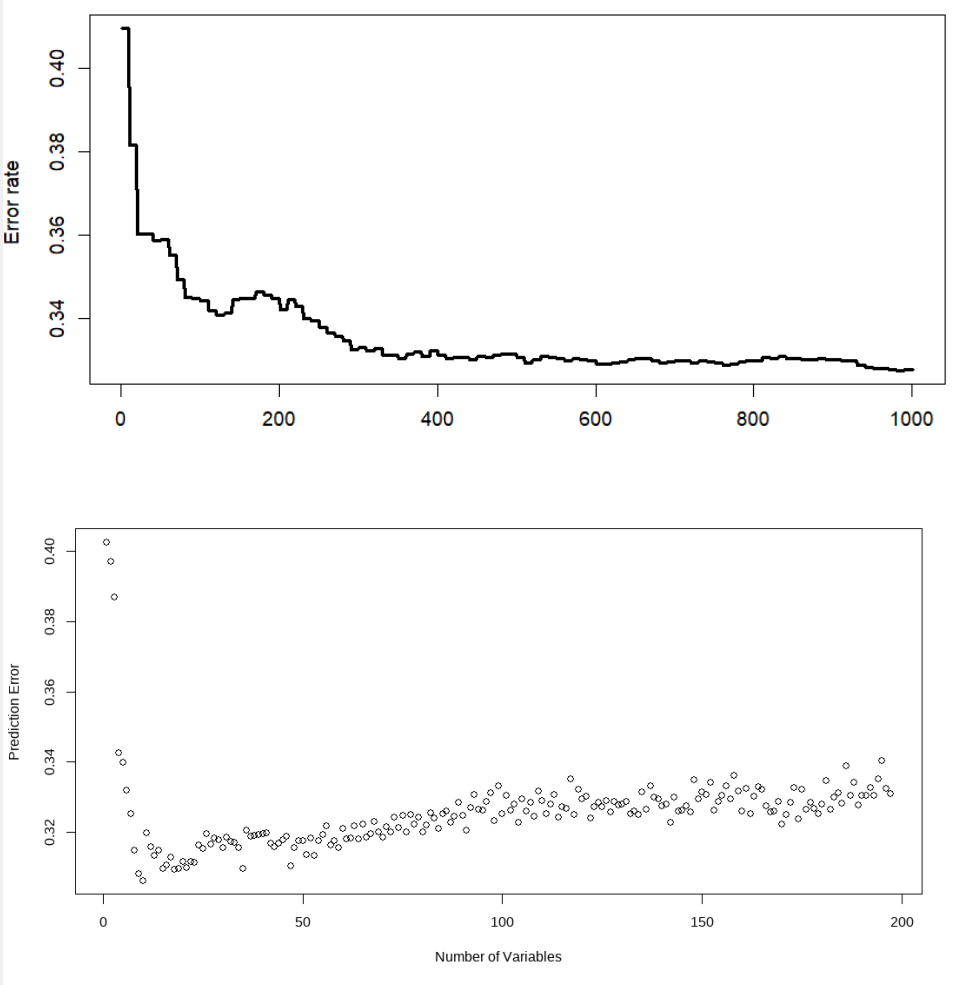
The models were scored based on Out of Bag (OOB) prediction error during model optimization.

Performance error responded as the number of variables falls

Performance error slowly improved until reaching a minimum using ten predictor variables, error: **0.3278.** The error sharply increased from that point.

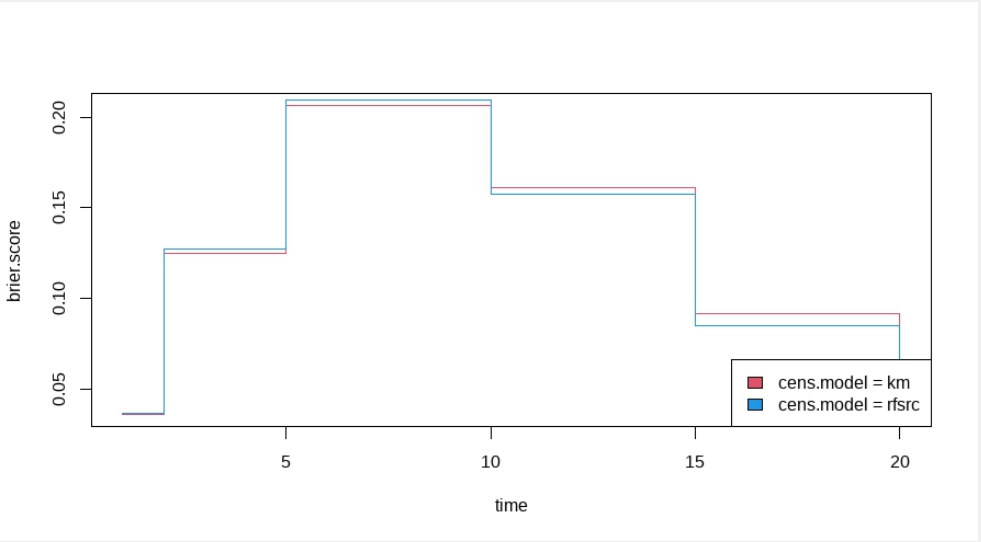
Applied to the test dataset (n=151X199) the performance error increased slightly to **0.369**.

- Shows good generalizability of the model.



Final Model & Model Fit

Brier score at specific times using Kaplan-Meier (KM-red) and Random Survival Forest (RSF-blue) censoring distribution estimators.



In survival analysis, the Brier Score -> accuracy of a predicted survival function at a given time.

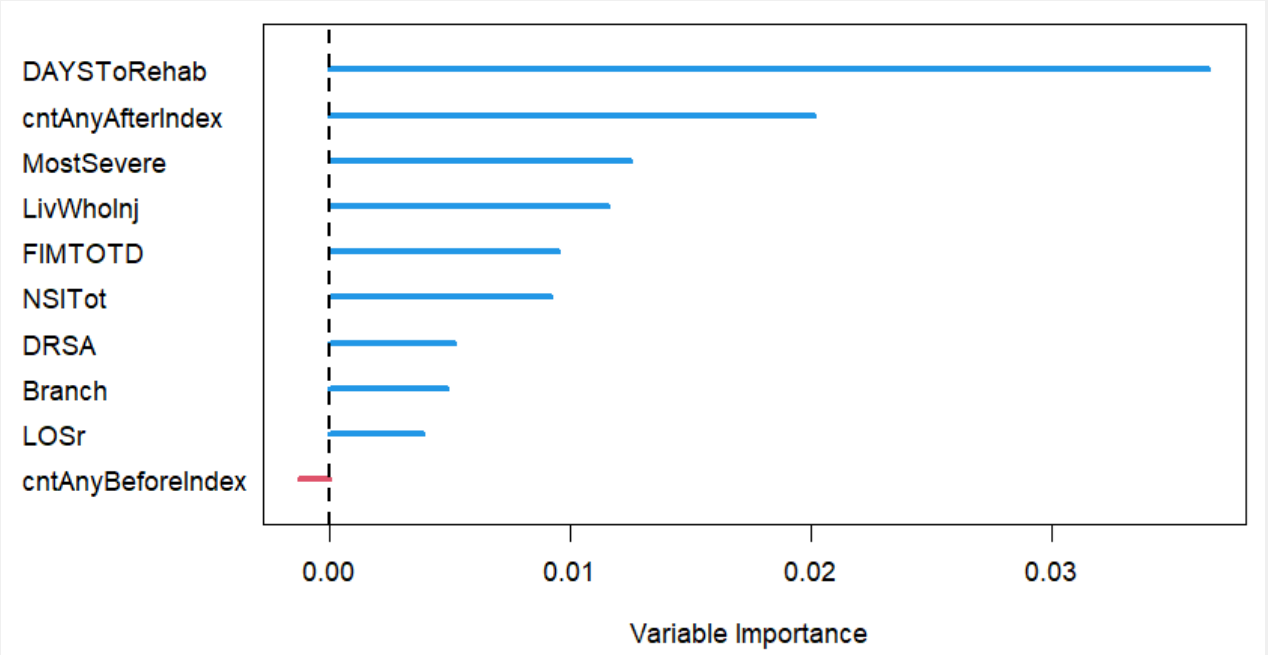
- Brier scores = 0 to 1 -> 0 indicates a perfect model.
- The integrated Brier score is the area under the RSF censoring distribution curve.
    - Can be used to assess the overall fit of a model over time.
- A model with an integrated Brier score below 0.25 is considered useful.
- **The integrated Brier score of the final model was 0.1231.**

Final Model & Variable Importance

Variable importance based on the OOB predictive error of the final model in the test set.

In Random Forest models, the importance of a variable reflects:

- Variable contribution to reducing uncertainty (error) in predictions
- Variables that are most influential in making accurate predictions, based on their effect on lowering model error.



Final Model & Variable Importance

Confidence intervals, standard errors, and exact variable importance values were calculated

Statistical Significance in traditional models is a hypothesis-testing concept that tells you whether an effect (e.g., the coefficient of a variable) is different from a null hypothesis in a statistically meaningful way.

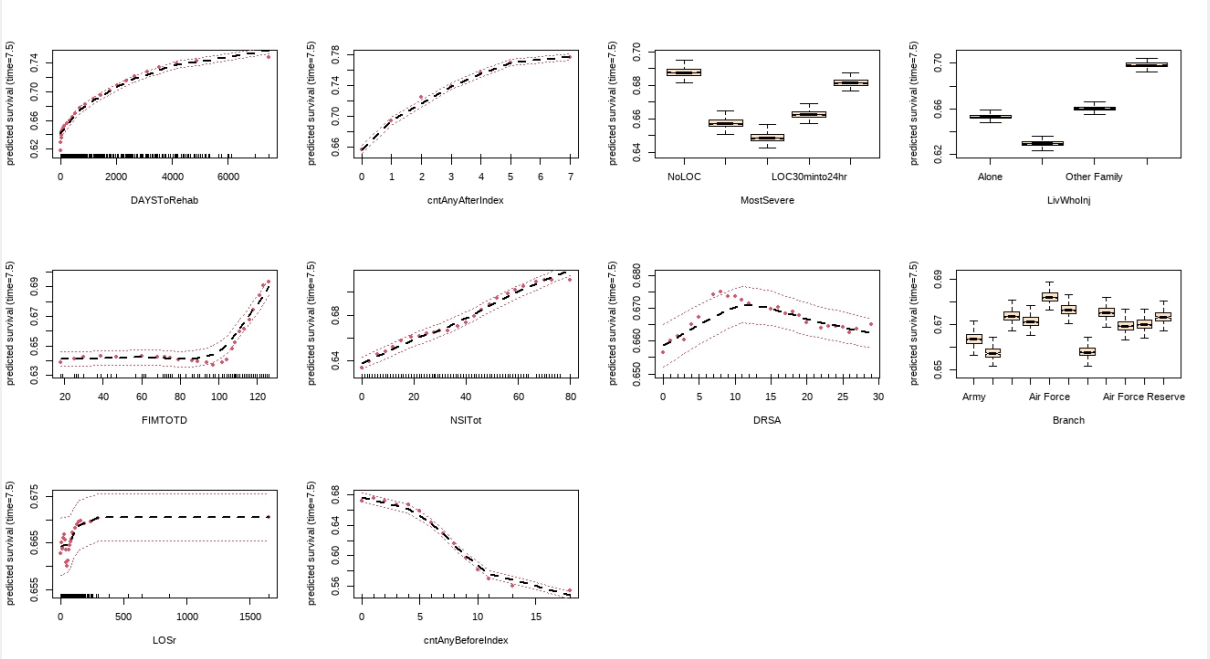
Based on this difference in meaning, the significance and the fact that one variable took on a negative importance value when applied to the test data should be acknowledged but not dismissed.

Variable	lower	mean	upper	p-value	signif
DAYSToRehab	0.0168	0.0365	0.0562	0.0001	TRUE
cntAnyAfterIndex	0.0122	0.0201	0.0280	0.0000	TRUE
MostSevere	0.0026	0.0125	0.0224	0.0069	TRUE
LivWhoInj	0.0009	0.0116	0.0223	0.0170	TRUE
FIMTOTD	0.0011	0.0095	0.0179	0.0133	TRUE
NSITot	0.0020	0.0092	0.0164	0.0060	TRUE
DRSA	-0.0012	0.0052	0.0116	0.0549	FALSE
Branch	-0.0021	0.0049	0.0119	0.0845	FALSE
LOSr	-0.0010	0.0039	0.0088	0.0586	FALSE
cntAnyBeforeIndex	-0.0049	-0.0012	0.0025	0.7419	FALSE

Final Model & Partial Dependence Plots (PDPs)

PDPs show the relationship between a feature and the predicted outcome.

- X-axis - the values of the feature that the PDP is being plotted.
- Y-axis shows the change in the predicted outcome as the feature value changes
    - Holding all other features constant.



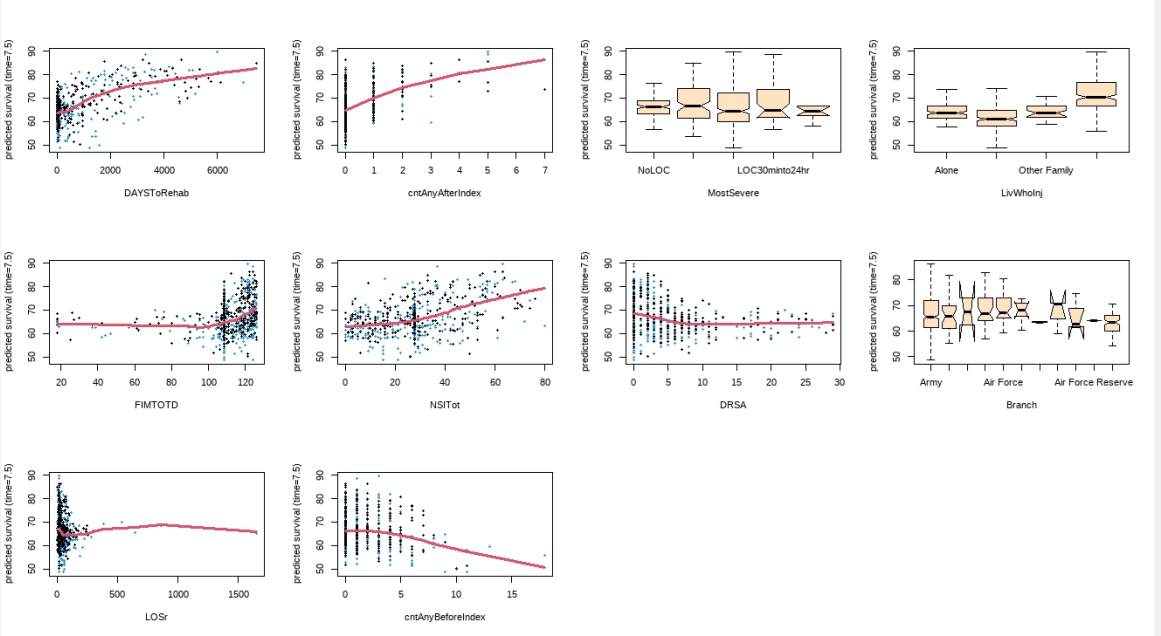
- **Upward slope:** indicating a positive relationship between the feature and the prediction.
    - As these features increase, the model predicts higher values for the probability of retention.
- **Downword slope:** indicates a negative relationship between the feature and the survival probability.
    - As the feature increases, the model predicts lower values for the probability of retention.

Final Model & Variable Predicted Survival Curves (PSC)

Variable PSCs - Visualize individual variable effects on the probability of retention over time.

Variable PSCs map the probability that an individual or a group will continue to participate in follow -ups beyond a certain time point.

The slope in PSCs is proportional to the variable's influence on the rate of LTFU over time.



The slope is key for interpreting predicted survival curves for individual variable effects, holding all other variables constant.

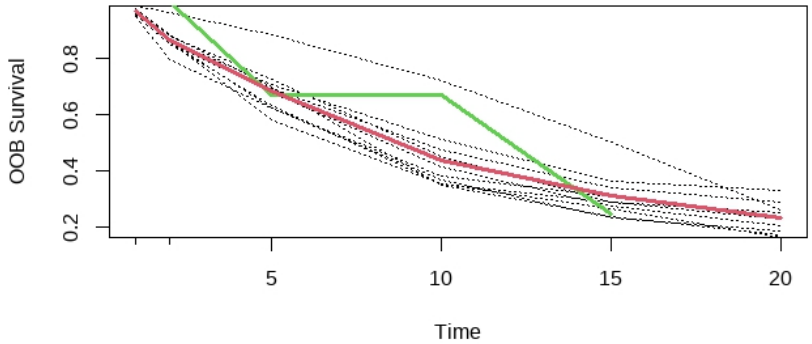
- **Upward slope:** indicates a higher rate of LTFU and a worse follow-up participation expectation.
- **Downward slope:** indicates a lower rate of LTFU and a better follow-up participation expectation.

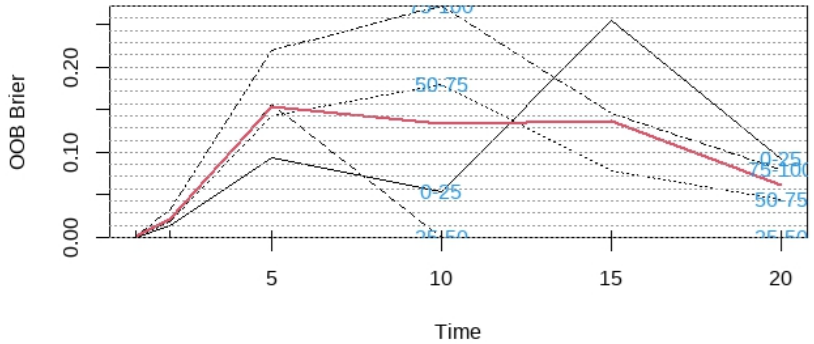
Can individual Predicted Survival Curves be plotted?

Predicted Survival Curve for Individuals

- Individual PSCs might provide insight into how survival probabilities change over time for a potential cohort.
- Allowing comparison of different individuals regarding time-to-LTFU
  - Note: outcomes ignore bias such as class differences.

Example individual and group predicted survival curves and OOB Brier scores:

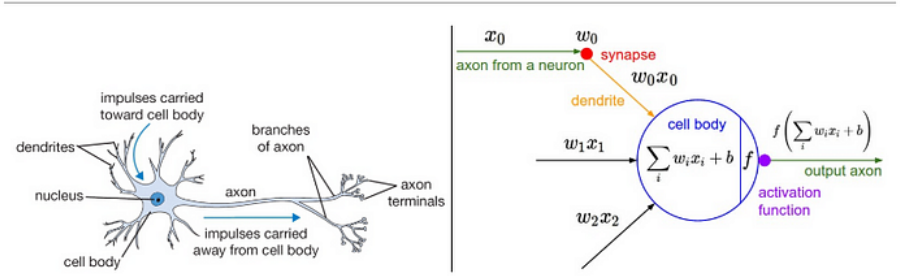




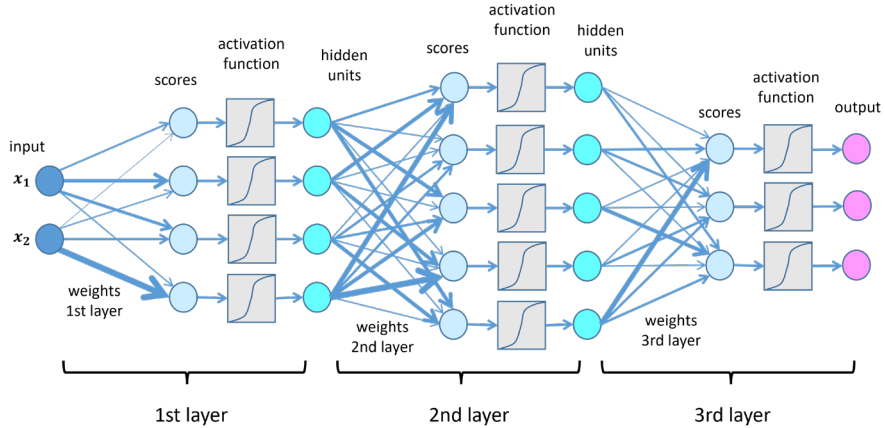
- Plots show the predicted probability of individual retention over time.
- Higher values indicate a greater likelihood of participation.

Example 2: Deep Learning and Neural Networks

Inspired by biological nervous systems



A cartoon drawing of a biological neuron (left) and its mathematical model (right).



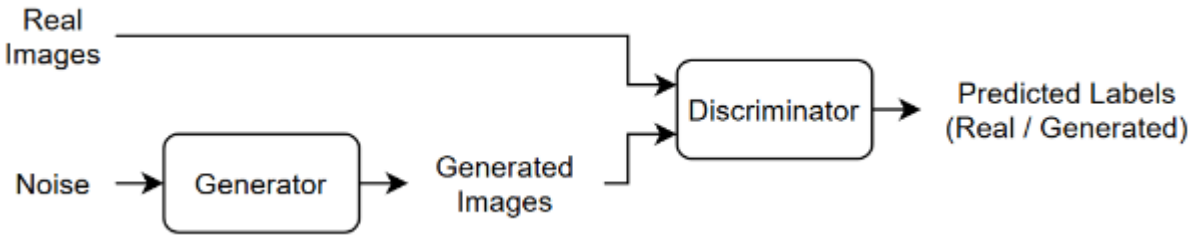
- Combines several processing layers using simple elements operating in parallel.
- Deep learning neural network consists of:
    - Input layer, multiple hidden layers, and an output layer.
- Each layer has many nodes, or “neurons”.
- Nodes in each layer use the outputs of all nodes in the previous layer as inputs.
    - Neurons interconnect with each other through the different layers.
- Each neuron is assigned an adjusted weight during the learning process.
    - Decreases or increases in the weight change the strength of that neuron’s signal.

Generative Adversarial Network (GAN)

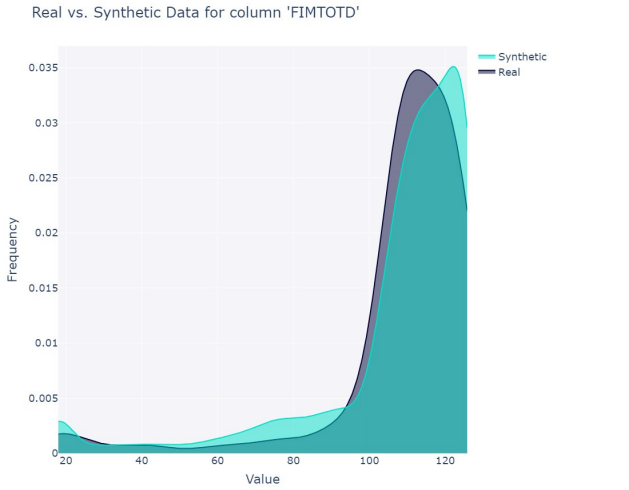
- A GAN is a type of deep learning generative AI that can create synthetic data with similar characteristics as the real data used for training.
- A GAN consists of two networks that train together:
  - Generator — Neural network generating data based on the training data.
  - Discriminator — Given batches of training data and generated data from the generator, this network attempts to classify the observations as "real" or "generated".

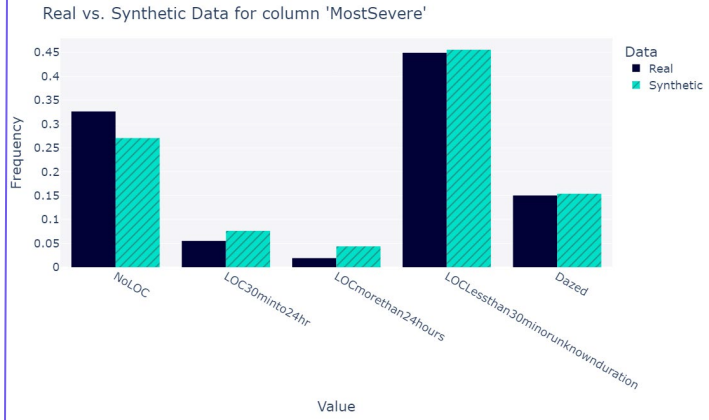
- Training a GAN: train both networks simultaneously - maximize the performance:
  - Train the generator to output data that "fools" the discriminator.
  - Train the discriminator to distinguish between real and generated data.

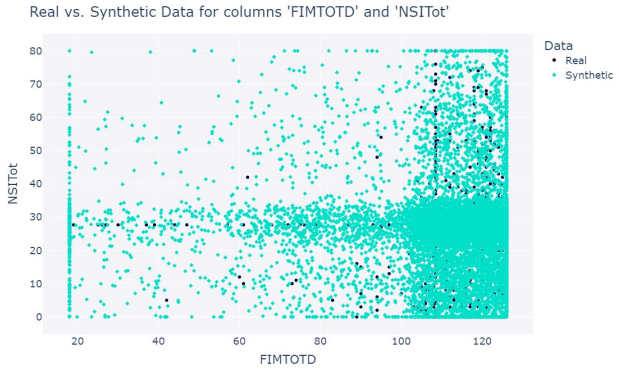
- Goal: a generator that creates realistic data and a discriminator that has learned strong feature representations of the characteristics of the training data.
  - Clip off the generator to make synthetic data.

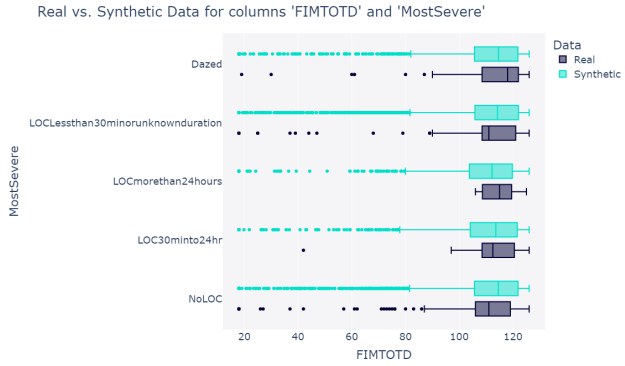


Generative AI (GAN), TBIMS Data









Name	Type	Size	Value
diagnostic	reports.singl…	1	DiagnosticReport objec…
discrete_columns	list	6	['LostReasonF', 'LivWh…
metadata	metadata.meta…	1	Metadata object of sdv…
real_data	DataFrame	(579, 12)	Column names: Followup…
synthesizer	single_table.…	1	CTGANSynthesizer objec…
synthetic_data	DataFrame	(10000, 12)	Column names: Followup…

Help | Variable Explorer | Plots | Files

Console 1/A

```
In [12]: metadata.visualize()
Out[12]:
```

RealData

FollowupPeriod : categorical
LostReasonF : categorical
LivWhoInj : categorical
DAYSToRehab : numerical
LOSr : numerical
FIMTOTD : numerical
DRSA : numerical
MostSevere : categorical
cntAnyBeforeIndex : numerica
cntAnyAfterIndex : categorica
NSITot : numerical
Branch : categorical

Primary key: None

```
In [13]: diagnostic = run_diagnostic(real_data=real_data,
synthetic_data=synthetic_data, metadata=metadata)
Generating report ...

(1/2) Evaluating Data Validity: |██████████| 12/12 [00:00<00:00,
704.93it/s]|
Data Validity Score: 100.0%

(2/2) Evaluating Data Structure: |██████████| 1/1 [00:00<00:00,
498.91it/s]|
Data Structure Score: 100.0%

Generating report ...

(1/2) Evaluating Column Shapes: |██████████| 12/12 [00:00<00:00,
113.21it/s]|
Column Shapes Score: 84.66%

(2/2) Evaluating Column Pair Trends: |██████████| 66/66
[00:00<00:00, 169.63it/s]|
Column Pair Trends Score: 82.65%

Overall Score (Average): 83.65%
```

Thank you.

Questions?

Research Support

The study was conducted with support from the VA TBI Model System, emphasizing collaborative research in understanding TBI.