WiM DRP Fall 2023

An Introduction to Explainable AI

Mentor: Trang Bui
Mentees: Sanika Poojary, Jaimee Yeung

Motivation

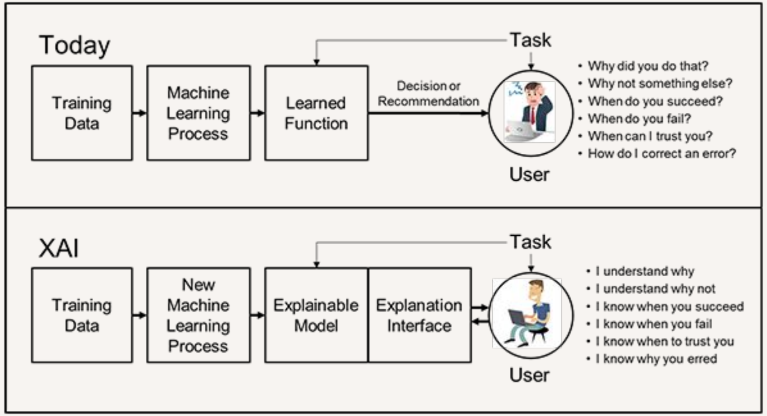
Motivation

What is Explainable AI?

- Explainable Artificial Intelligence (AI) is the ability for artificial intelligence systems to provide understandable explanations for their decisions, recommendations, or predictions

Why do we need Explainable AI?

- AI is becoming a crucial part of our lives
- Models are mostly “black-box” (ie. user does not know how the model actually works)
- Uses of Explainable AI:
    - Accountability & Responsibility
    - Transparency builds trust among users
    - Legal & Ethical Compliance
    - Bias Detection & Mitigation
    - Facilitating user understanding
    - Advancing research & collaboration between humans and AI



Dataset: Early Classification of Diabetes

According to the World Health Organization (WHO), diabetes is one of the *fastest growing chronic diseases*. Early detection is essential in helping diagnose and treat diabetic patients.

About the Dataset

- 520 observations (ie. patients) with 16 characteristics/symptoms (ex. Age, gender, etc.)
- Data was collected through questionnaires and diagnosis results from the patients in the Sylhet Diabetes Hospital in Bangladesh

Objective

- With the use of ML models, **predict the diagnosis of diabetes** using patient's profile + characteristics
- Identify the characteristics/symptoms that have the highest contribution to the diagnosis of diabetes through multiple machine learning models

Machine Learning Models

What are Machine Learning Models?

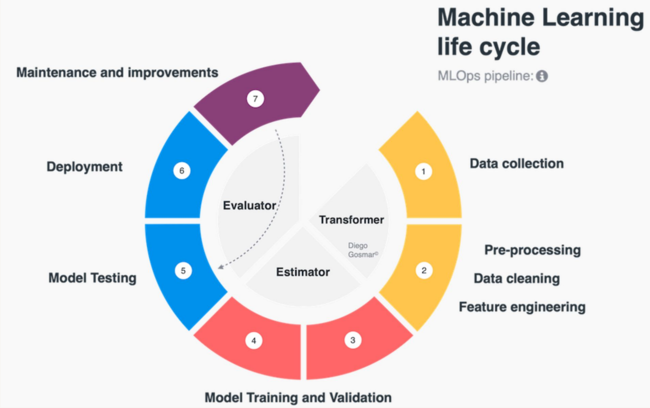
- **What is the purpose of ML models?**
  - To enable data-based learning, reasoning, and decision making via statistical models and computational algorithms
  - Some other key purposes include: Classification, Regression, Clustering, etc.

- **How do ML models work?**
  - Use mathematical models to generalize from input features and corresponding output labels in training data, allowing them to make predictions on new data
  - The algorithms **iteratively adjust their internal parameters** during training to minimize the difference between predicted and actual outcomes, enabling them to generalize and perform well on diverse datasets

What are Machine Learning Models? (cont'd)

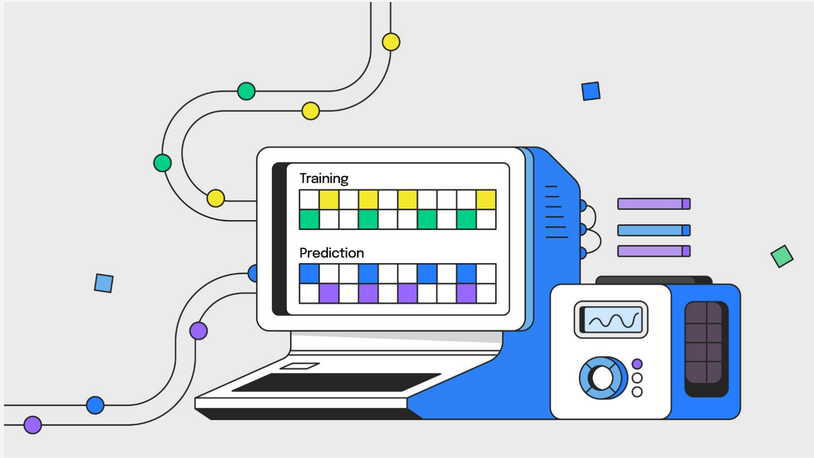
- **Examples of ML Models:**
  - Linear Regression, Logistic Regression, Decision Trees, Random Forest, K-Nearest Neighbours, Naive Bayes, Neural Networks and many more!

- **How to evaluate ML Models**
  - The choice of evaluation metrics depends on the type of problem (ex. Classification, Regression, Clustering) and the specific goal of the model
  - Regression Models: Mean Absolute Error (MAE), Mean Squared Error (MSE), R-Squared, etc.
  - Classification Models: Precision, Accuracy, Confusion Matrix, F1 Score, etc.
  - Compare the model's performance on the training & test sets.
  - We will be focusing on using the **Mean Squared Error (MSE)**



What are Machine Learning Models? (cont'd)

- **How to evaluate ML Models**
  - First, we split the dataset into the training and test data
  - **Training Set:** The data that is used to train the ML model.
    - This data contains true labels.
  - **Test Set:** The data that is used to evaluate the performance of the ML models (ie. using MSE).
    - The true labels are removed and utilize the ML models to make predictions based on the features in the test data.
    - We then compare the predictions with the true labels to evaluate the ML models

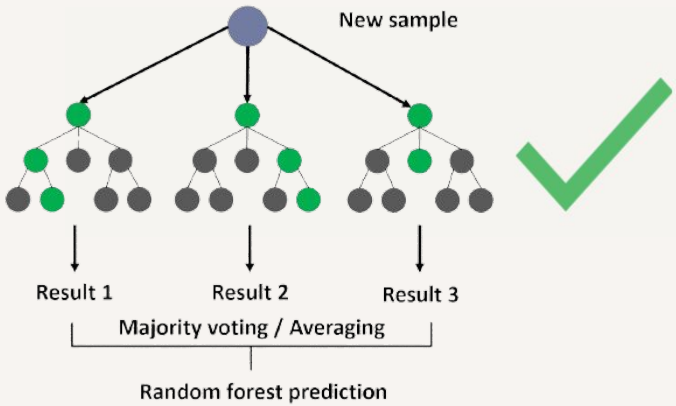


Random Forest

- ***Random forest*** is a ML algorithm that uses the results of multiple “decision trees ” to create a single result.
    - What is a “**decision tree**”?
        - A decision tree is a flowchart-like model that represents a decision-making process, where each node denotes result of a decision.
    - Why is it **random**?
        - The data (including the observations and features) used to train the model for each decision tree is randomly selected

- Each decision tree in the forest makes a prediction

- The individual predictions of the ensemble of trees are combined to create a single result
    - This process is also called “**ensemble learning**”

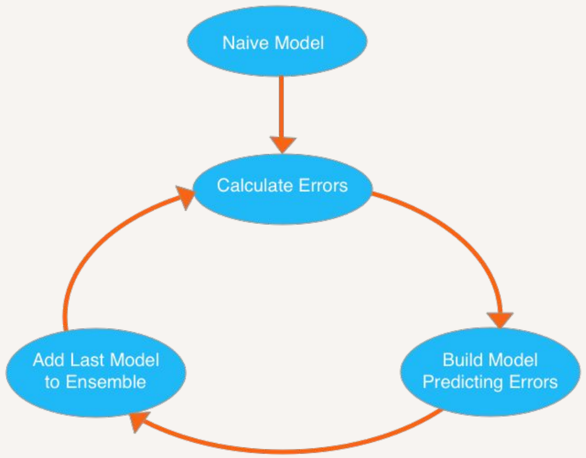


XGBoost

- **What is an "Ensemble Learning Algorithm"?**
  - A technique of combining the predictions of multiple machine learning models to improve overall performance & robustness
  - Used to reduce overfitting, enhance generalization, and increase predictive accuracy
  - Ex. Random Forest, XGBoost

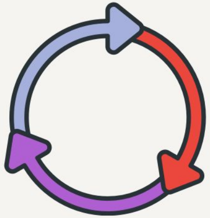
- **XGBoost (eXtreme Gradient Boosting)** is an ML algorithm that is an implementation of *gradient boosted decision trees*, designed for speed and performance.
  - Works by combining the predictions of multiple weak models (usually decision trees) additively
  - New trees are built based on the performance of old trees

- There is no straightforward way to interpret the XGBoost or Random Forest model because they are "Black-Box Models"



Permutation Importance

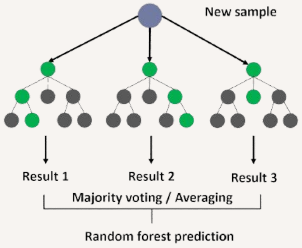
Permutation Importance

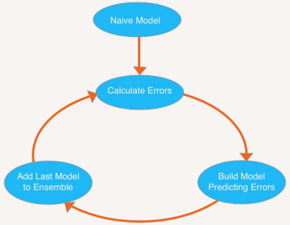


- **What is *Permutation Importance?***
- Permutation feature importance measures the increase in the prediction error of the model after permuting the feature values.
    - This process breaks the relationship between the feature and the true outcome.
        - i.e. permutation feature importance takes into account both the main feature effect and the interaction effects on model performance

- **How does it work?**
    - To calculate the importance score for each feature:
        - In the test dataset, we permute the observations (rows) within the feature while keeping the other features the same
        - Use the machine learning model to make predictions based on the new test dataset
        - Then we evaluate the prediction error and compare it with the prediction error made using the original dataset

Permutation Importance





Input

Machine Learning Model

Process

Permute Each Feature

Output

Feature Importance Scores

Advantages & Disadvantages

- **Advantages**
  - **Easy Interpretation**
    - Feature importance is the increase in model error when the feature's information is destroyed
  - **Comparable**
    - Feature importance measurements can be compared across different problems
  - **Does Not Require Model Retraining**
    - Permuting a feature can save a lot of time: there importance methods go through the process of deleting a feature, retraining the model and then comparing the model error

- **Disadvantages**
  - **Requires Access to the True Outcome**
    - Permutation importance cannot be calculated without the true results of a dataset
  - **Results May Vary**
    - The permutation feature importance is determined by shuffling the feature, which adds randomness to the measurement

Data Analysis

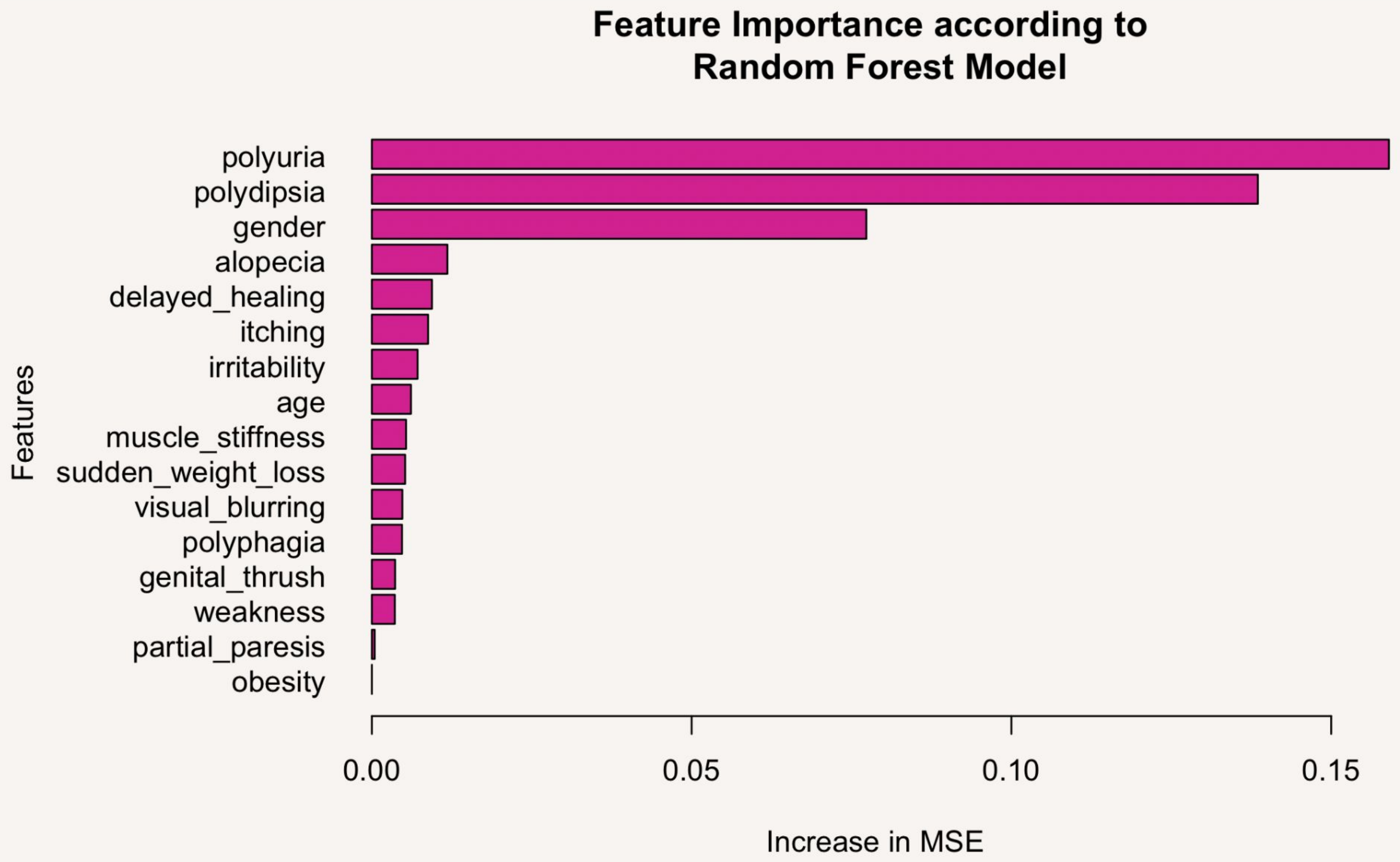
Overview

- Used R to perform our analyses
- 80% of the data is randomly selected as the training set, and the remaining 20% is the test set
- Used Random Forest, and XGBoost to predict the diagnosis of diabetes based on the characteristics and symptoms of the patient
- Mean Squared Error (MSE): Measures how well a predicted value matches some truth value

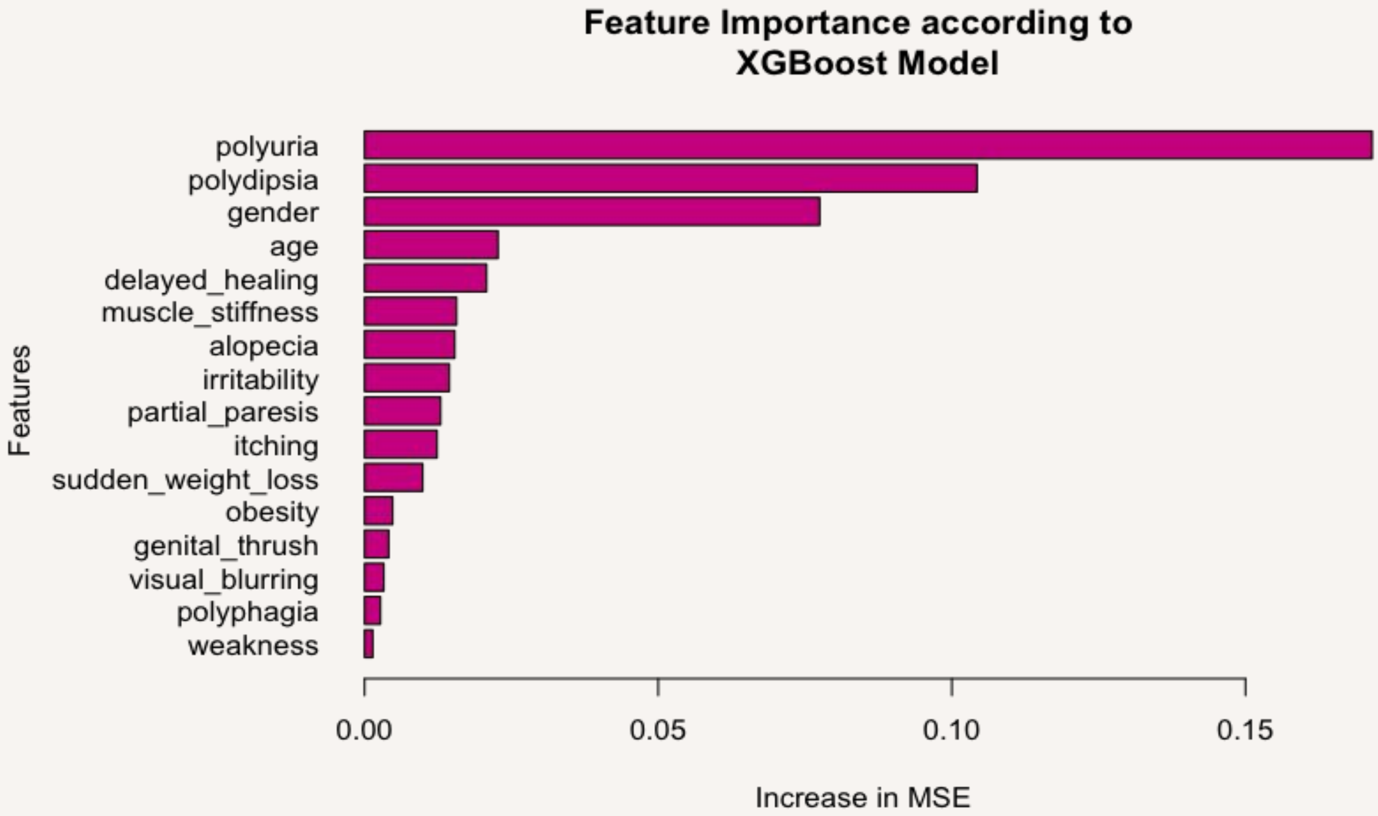
$$\mathrm{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}\left(\overset{\text{Error}}{\boxed{Y_i - \hat{Y}_i}}\right)^{\overset{\text{Squared}}{\boxed{2}}}$$

Model	Training Errors	Test Errors
Random Forest	0	0.01923077
XGBoost	5.206861e-06	0.03179245

Random Forest



XGBoost


Feature Importance according to
XGBoost Model
polyuria
polydipsia
gender
age
delayed_healing
muscle_stiffness
alopecia
irritability
partial_paresis
itching
sudden_weight_loss
obesity
genital_thrush
visual_blurring
polyphagia
weakness
Features
0.00
0.05
0.10
0.15
Increase in MSE

Importance Analysis

From both the Random Forest and XGBoost models we saw that polyuria, polydipsia and gender were the most important features in diabetes diagnosis:

- **Polyuria:**
  - Whether a patient experienced excessive urination or not
    - Patients with polyuria are more likely to have diabetes than those who do not
- **Polydipsia:**
  - Whether the patient experienced excessive thirst/excessive drinking or not
    - Patients with polydipsia are more likely to have diabetes than those who do not
- **Gender**
  - Females are more likely to have diabetes than males

Thank You!