{ "cells": [{ "cell_type": "markdown", "metadata": {}, "source": ["# Predicting PIM3 Risk of Mortality with Fewer Variables using Supervised Machine Learning "] }, { "cell_type": "code", "execution_count": 1, "metadata": {}, "outputs": [{ "name": "stdout", "output_type": "stream", "text": ["Sklearn version: 0.24.2\n", "LightGBM version: 2.3.1\n", "Pandas version: 0.25.1\n", "Numpy version: 1.16.5\n", "SHAP version: 0.34.0\n"] }], "source": ["###Import Libraries\n", "import pandas as pd\n", "import numpy as np\n", "import seaborn as sns\n", "import sklearn\n", "from scipy import stats\n", "import matplotlib.pyplot as plt\n", "%matplotlib inline\n", "import graphviz\n", "import shap\n", "import lightgbm as lgb\n", "from lightgbm import LGBMModel\n", "from lightgbm import LGBMClassifier\n", "from lightgbm import LGBMRegressor\n", "from sklearn import ensemble\n", "from sklearn.ensemble import GradientBoostingClassifier\n", "import math\n", "from sklearn import preprocessing\n", "from sklearn.model_selection import GridSearchCV\n", "from sklearn.model_selection import ParameterGrid\n", "\n", "from scipy.stats import uniform as sp_rand\n", "from sklearn.model_selection import RandomizedSearchCV\n", "from sklearn import model_selection\n", "from sklearn.linear_model import LogisticRegression\n", "from sklearn.model_selection import train_test_split\n", "from sklearn import preprocessing\n", "from sklearn.preprocessing import LabelEncoder\n", "from sklearn.preprocessing import OneHotEncoder\n", "from sklearn.metrics import mean_absolute_error\n", "from sklearn.metrics import mean_squared_error, r2_score\n", "from sklearn.metrics import accuracy_score\n", "from sklearn.metrics import confusion_matrix\n", "from sklearn.model_selection import KFold\n", "from sklearn.model_selection import StratifiedKFold\n", "from sklearn.model_selection import cross_val_predict\n", "from sklearn.model_selection import cross_val_score\n", "from sklearn.metrics import classification_report\n", "from sklearn.metrics import roc_auc_score\n", "from sklearn.metrics import roc_curve\n", "from sklearn.metrics import precision_recall_curve\n", "from sklearn.metrics import f1_score\n", "from sklearn.metrics import auc\n", "from sklearn.metrics import average_precision_score\n", "from sklearn.metrics import precision_score\n", "from sklearn.metrics import recall_score\n", "from sklearn.metrics import brier_score_loss\n", "\n", "pd.options.display.max_columns = None\n", "pd.options.display.max_rows = None\n", "\n", "print ('Sklearn version: '+sklearn.__version__)\n", "print('LightGBM version: '+lgb.__version__)\n", "print('Pandas version: '+pd.__version__)\n", "print('Numpy version: '+np.__version__)\n", "print('SHAP version: '+shap.__version__)"] }, { "cell_type": "code", "execution_count": 2, "metadata": {}, "outputs": [{ "name": "stdout", "output_type": "stream", "text": ["(13324, 154)\n"] }, { "data": { "text/html": ["
\n", "