# FDMG: An Interactive Integrated Analytics Tool for Financial Data

Qingjian Li, Jiaji Lu, Chunquan Wu, Chonglong Yu

## Motivation and Introduction

- Financial analysts might not have an easy-to-use tool to visualize and analyze a large amount of alternative datasets.
- The traditional solution involves extensive programming, which might not be intuitive to users not familiar/comfortable with the programming environment, including senior management, traders, traditional financial analyst.
- It might be difficult for financial analysts or others that don't have much knowledge of machine learning to interpret the results of various machine learning models and pick the right model for decision making.

A web-based interactive integrated analytics tool (FDMG: stands for Financial Data Modeling Guru) specialized in predictive analytics and visualization of financial data via cutting edge machine learning algorithms.
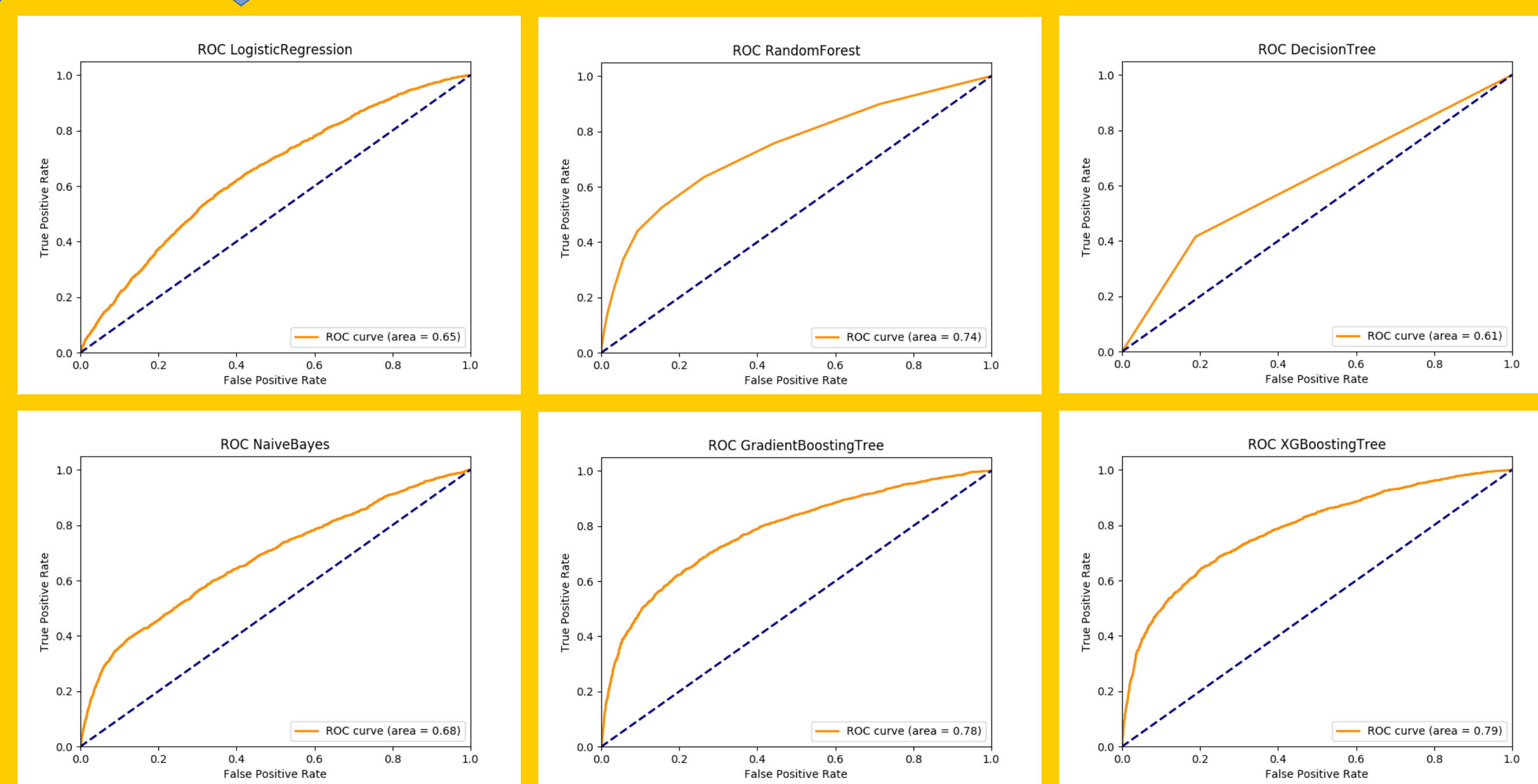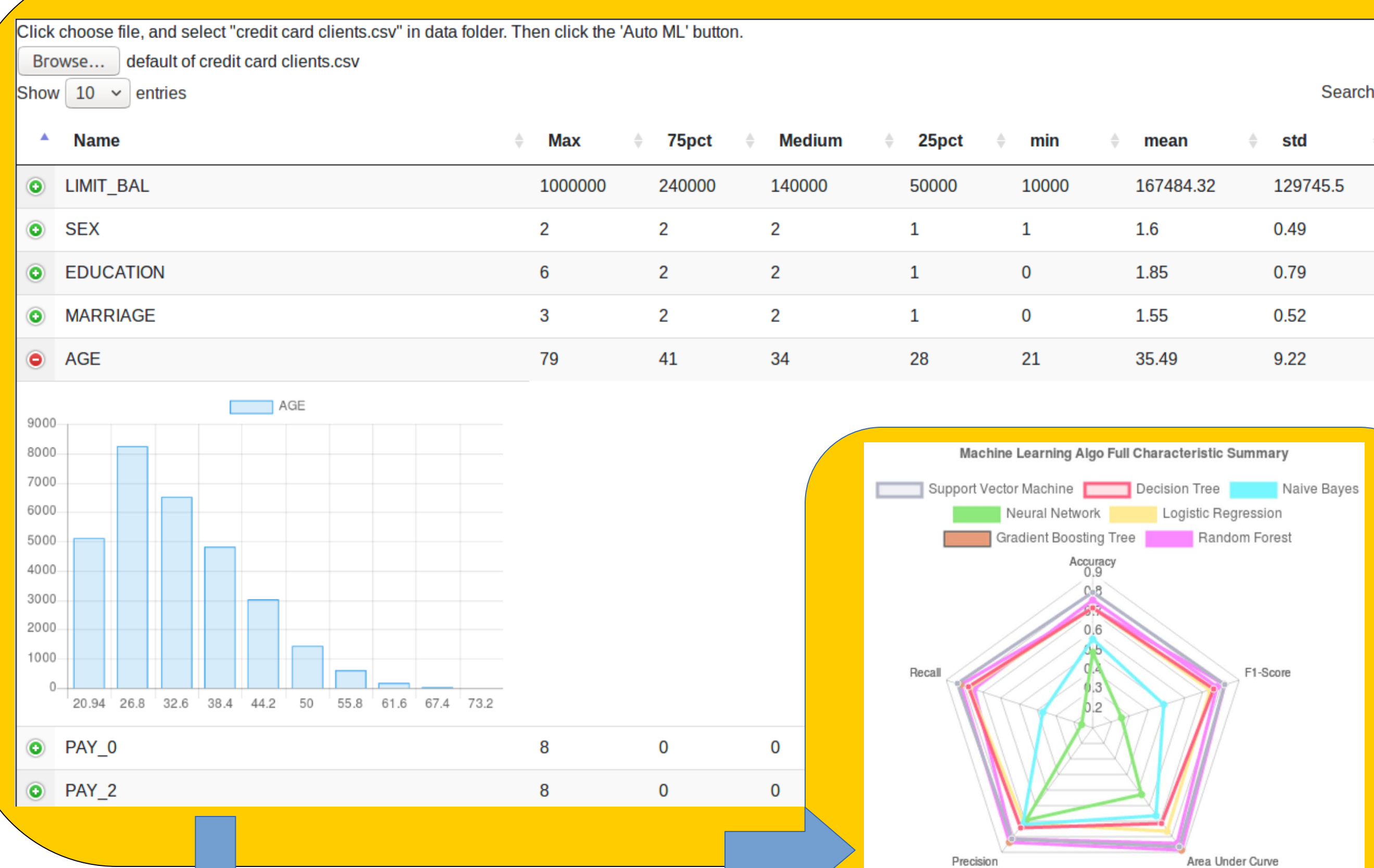
## Features

- Provide a user-friendly GUI to visualize the data through histogram and the key statistics computed.
- Provide an automated ML framework for the user to determine the best ML models to use, and automatically generate model diagnostics.

## Approach

### Data Visualization

**User upload their dataset**



### Classification Algorithm

**User select a column to predict**

**A series of Machine Learning algorithms are executed at back end**

| Logistic Regression | Random Forest |
| Gradient/XG Boosting Tree | Naive Bayes |
| Support Vector Machine | Decision Tree |

**Evaluate a set of metrics: out-of-sample accuracy, ROC, log-loss, etc**

**Models ranked by a combination of metrics**

**Rank of models and prediction values are presented to user**

### Tech Stack
python, Flask, D3, JavaScript, jQuery.JS and AJAX, scikit learn, App Engine

## Data

We use a Credit Card Default Dataset as an experiment dataset (https://archive.ics.uci.edu/ml/machine-learning-databases/00350/):
- A total of 30000 rows and 24 columns.
- Last column indicates credit card default or not.
- Features include amount of given credit, gender, education, marital status, age, etc.
- We aim to use FDMG to predict default probability given the features.

## Evaluation

- We use out-of-sample accuracy, ROC (AUC), confusion matrix and Log loss to assess performance of candidate models.
- XG Boosting Tree is recommended by FDMG as the top ML model for the experiment dataset.
- XG Boosting Tree performs much better than linear models that are widely used in the finance industry.
- Our result generally agrees with results of Yeh and Lien, ESA, 2009, except that several recent models are included in our experiment.

To try FDMG, please visit https://fdmg-team35.appspot.com/. Contact information: qli387@gatech.edu, jlu314@gatech.edu, gth752t@gatech.edu, cyu342@gatech.edu.