SAS®
GLOBAL
FORUM
2020

MARCH 29 – APRIL 1
WASHINGTON, DC

USERS PROGRAM

SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Abstract
Introduction
Methods
Results
Conclusion

Biography

Raphael Lima is a Brazilian mathmatician and Data Scientist, specializing in unsupervised learning and segmentation analysis. Over 10 years of professional experience in demanding positions, regarding both applying analytics to decision making and data Science skills. Mathematics degree from Universidade Federal de São Carlos(UFSCAR), DataScience MBA candidate from Instituto de Ensino e Pesquisa(INSPER) and Data Science Professional Certified Using SAS 9. Strong Backgroud in SAS, R and Python, advanced analytics and machine learning methods.

Raphael Lima

Abstract

Making good recommendations are essential for the retail and wholesale markets, and in many cases these suggestions become a competitive differential when aligned with marketing and sales campaigns. Empowered by those market trends big companies like Netflix in their streaming platform, or giants like Amazon and Airbnb are working hard to improve their recommendations always seeking for customer satisfaction.

A Recommender System(RS) is a software with models that provide items suggestions for its user appreciation, like thousands of Spotify® subscribers, we are used to get a good custom playlist recommendation every week, but every song wrongly recommended, makes us wonder how to make better suggestions, so we decided to consume our data from Spotify API and recommend our own songs. In this scenario, we will develop two Recommender Systems, first one using SAS® Enterprise Miner and another one using Python-Scikit-Learn, and evaluate the accuracy of both modelling tools, and the results were amazing!
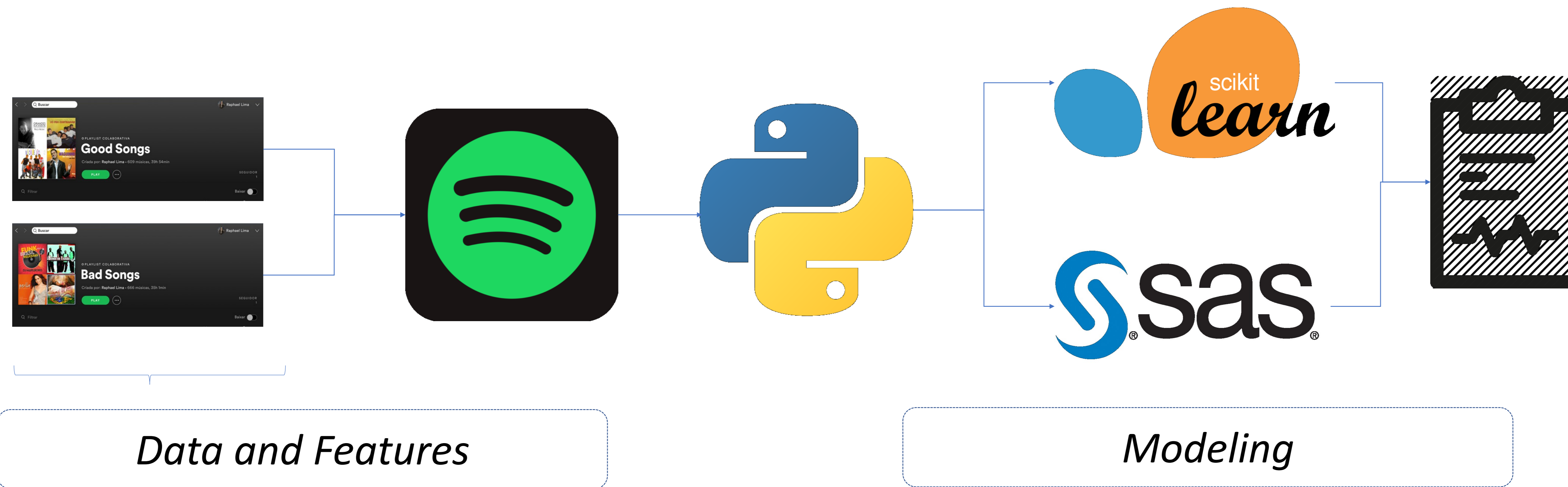
SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Intro

We decided to compare SAS Enterprise Miner and Python Scikit-learn as they are excellent modeling tool widely used in Data Science and Statistics fields. We keep up with the growing expansion of open source tools in novice programmers, while the SAS tool, specifically SAS® Enterprise Miner, has traditionally been used in the industry by professionals with a broader market experience. However, besides being an open source tool, what advantages does Python have compared to SAS in decision model building?
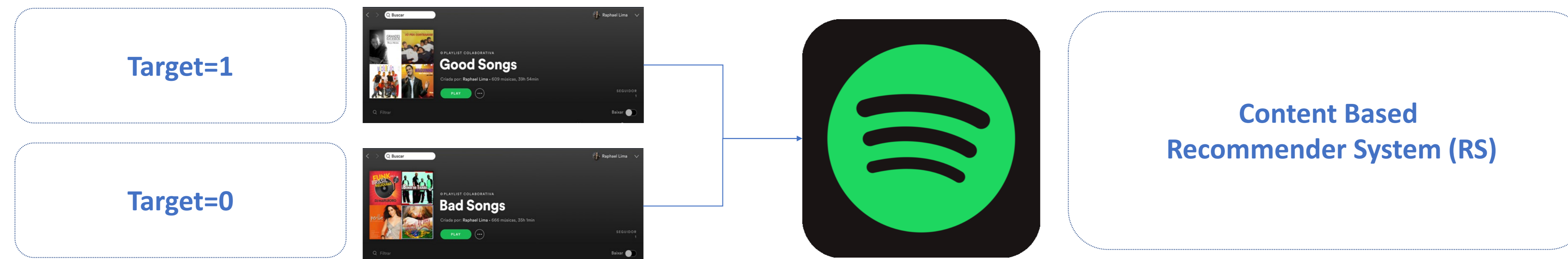
Research - Workflow

Data and Features

Modeling

SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Data

Target=1

Target=0

Content Based
Recommender System (RS)

Source

Features

Accousticness	Valence
Duration	Energy
Loudness	Speechiness
Tempo	Instrumentalness
Danceability	Key
Liveness	

SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Abstract
Introduction
Methods
Results
Conclusion

SAS Enterprise Miner - Flux
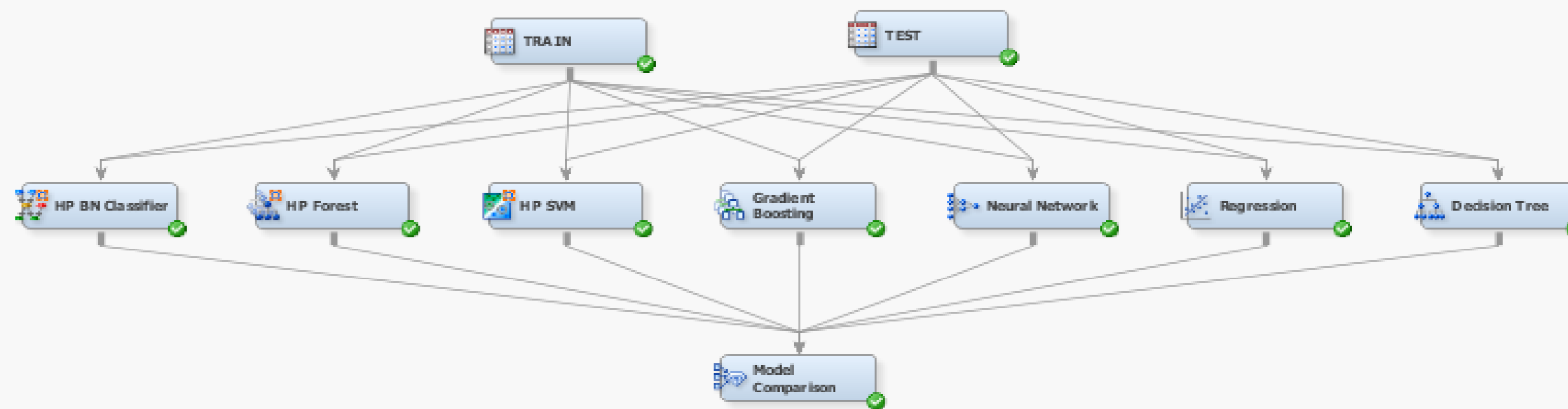
Model Comparison
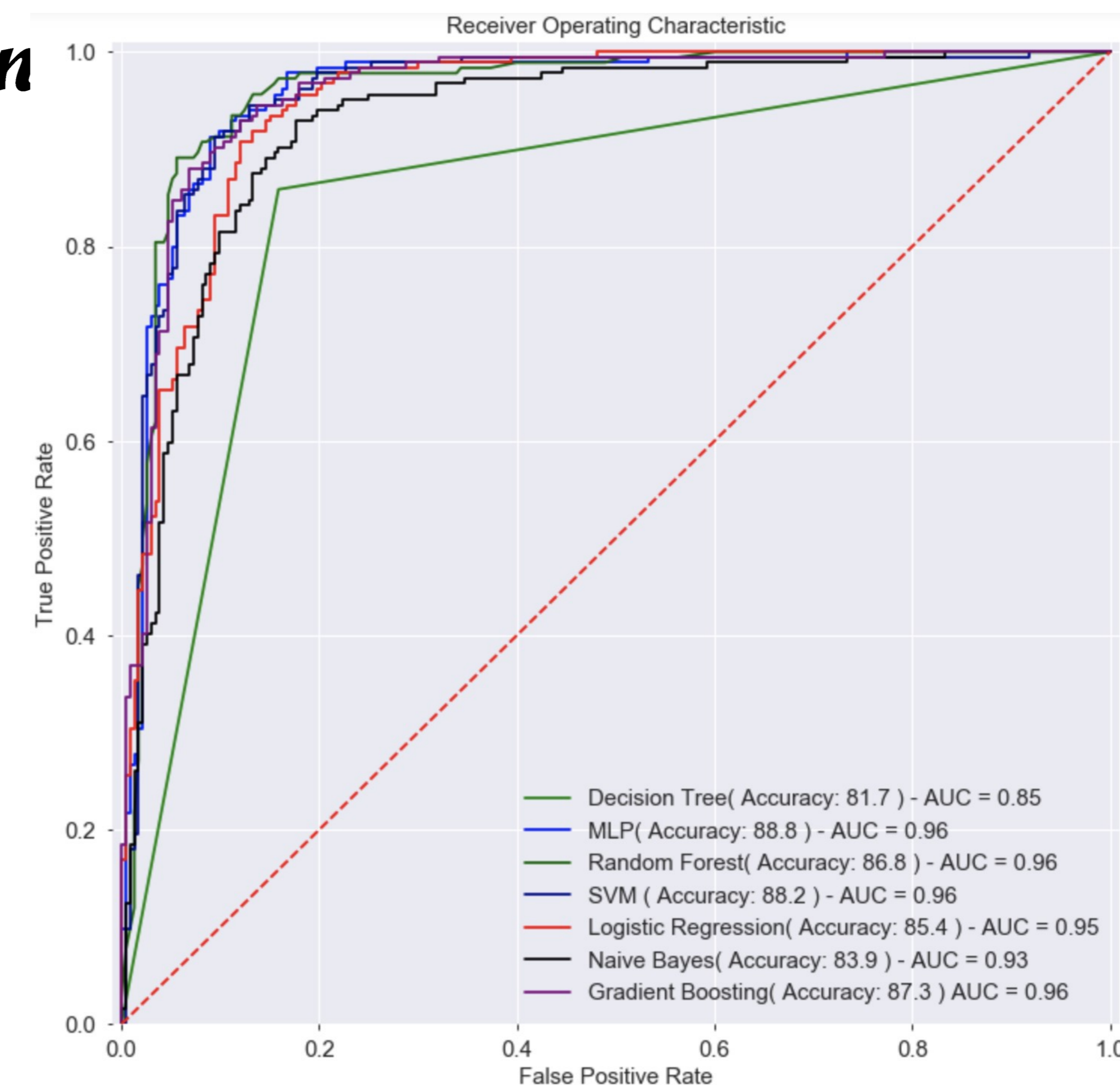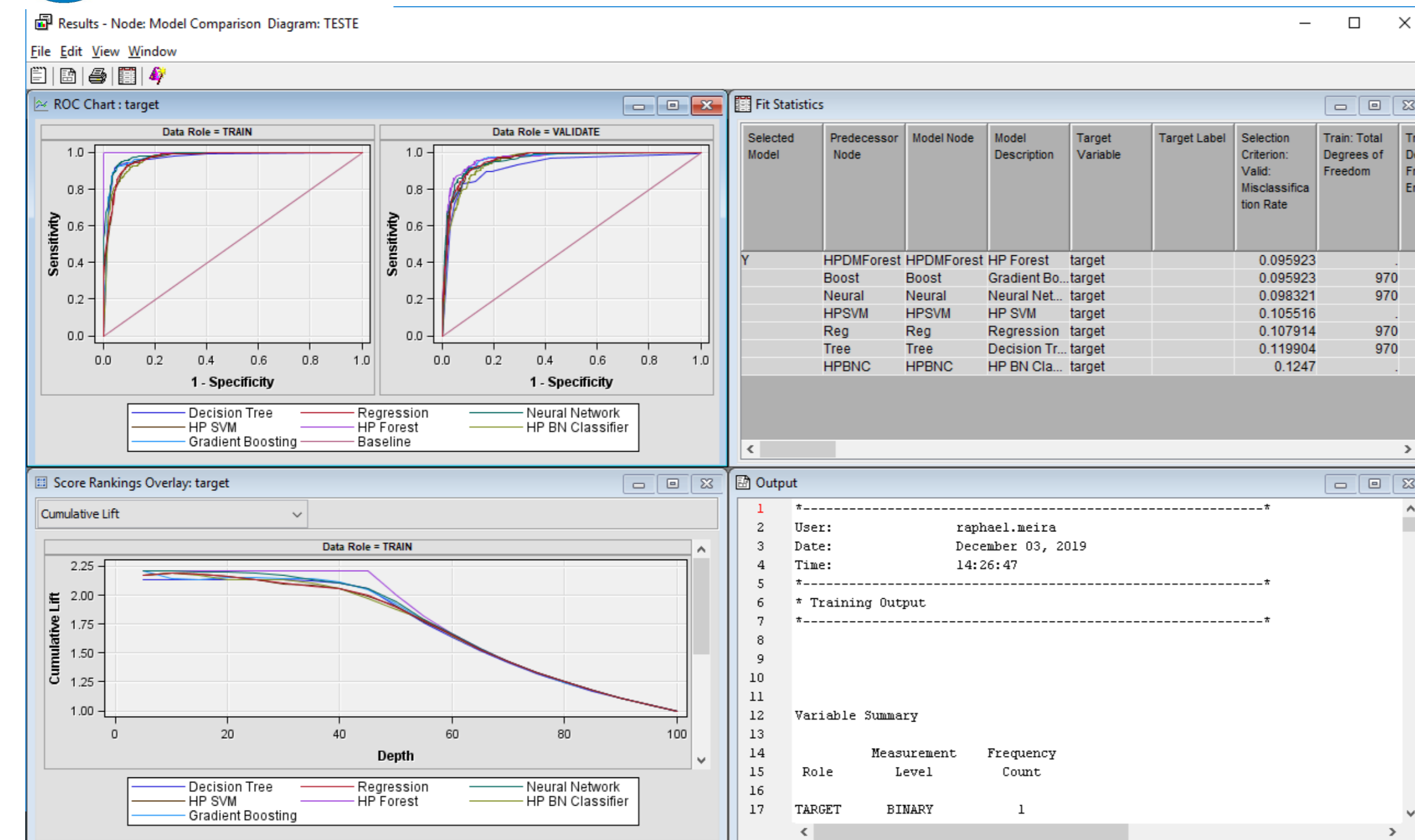
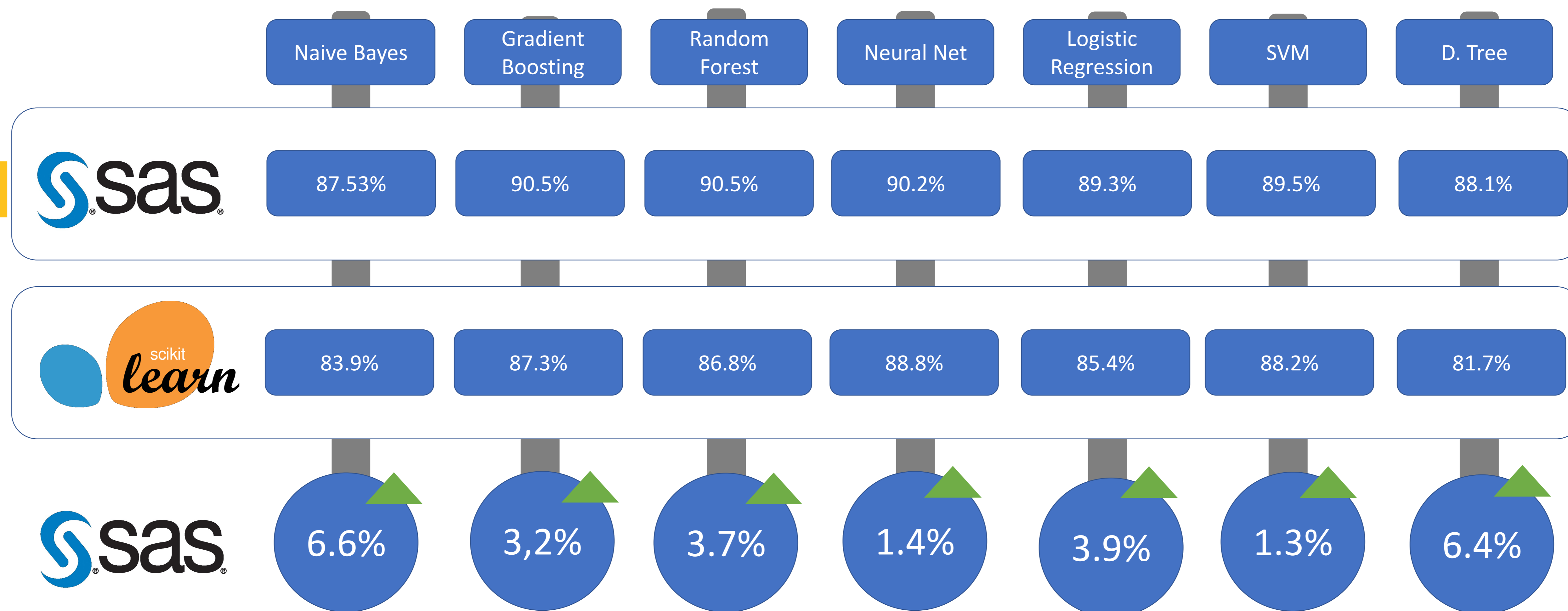# SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Abstract
Introduction
Methods
Results
Conclusion

Accuracy Score

	Naive Bayes	Gradient Boosting	Random Forest	Neural Net	Logistic Regression	SVM	D. Tree
SAS	87.53%	90.5%	90.5%	90.2%	89.3%	89.5%	88.1%
scikit learn	83.9%	87.3%	86.8%	88.8%	85.4%	88.2%	81.7%
SAS	6.6%	3,2%	3.7%	1.4%	3.9%	1.3%	6.4%

SAS® Enterprise Miner™ proved superior to Scikit Learn in Accuracy Score(Lower Misclassification Rate) when we used the default configuration of its in comparison to default configuration of Scikit Learn Library.
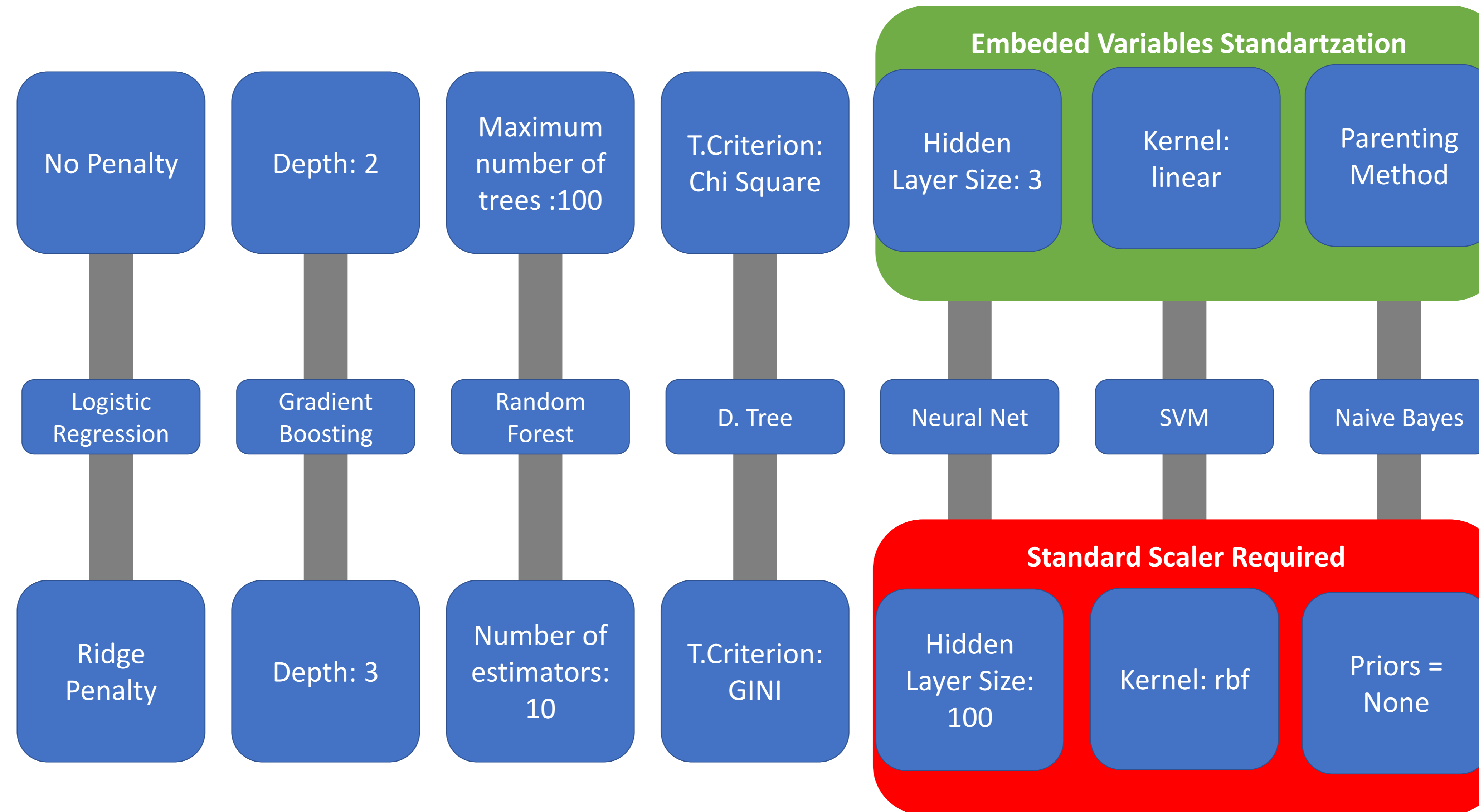
SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Default Comparison(Documentation)

SAS

No Penalty	Depth: 2	Maximum number of trees :100	T.Criterion: Chi Square

Embeded Variables Standartzation

Hidden Layer Size: 3	Kernel: linear	Parenting Method

Logistic Regression	Gradient Boosting	Random Forest	D. Tree	Neural Net	SVM	Naive Bayes

scikit learn

Ridge Penalty	Depth: 3	Number of estimators: 10	T.Criterion: GINI

Standard Scaler Required

Hidden Layer Size: 100	Kernel: rbf	Priors = None

SAS® Enterprise Miner^{TM} vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Conclusion

In this case study, comparing classification methods , we concluded that with SAS Enterprise Miner, it was possible to built models with higher accuracy and low missclassification rate under validation sample than Scikit learn in python considering the default settings of each method in both solutions.

Acknowledgments

The author is grateful to SAS Customer Care team – Camila Reis and Fernanda Knopki for their valuable assistence in this poster.

References

Hastie, T., Tibshirani, R. & Friedman, J.H. (2017) "The Elements of Statistical Learning ", Springer,
Ricci, F., Rokach, L. Shapira, B. & Kantor , P,B. (2011) "Recommender Systems Handbook", Springer, ISBN: 0387952845.
Spotify for Developers,(Available in 09/30/2019). http://developer.spotify.com/documentation/
SAS® Enterprise Miner^{TM} 14.3: Reference Help ,(09/30/2019).
https://documentation.sas.com/?cdcId=emlearncdc&cdcVersion=1.0&docsetId=emex&docsetTarget=titlepage.htm&locale=pt-BR

Abstract
Introduction
Methods
Results
Conclusion

Raphael Lima
Data Science & Analytics

SAS®
GLOBAL FORUM 2020

USERS PROGRAM

MARCH 29 - APRIL 1 | WASHINGTON, DC | #SASGF