SAS®
GLOBAL FORUM
2020
MARCH 29 - APRIL 1
WASHINGTON, DC
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!
We decided to compare SAS Enterprise Miner and Python Scikit-learn as they are excellent modeling tools 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?
SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

Data

- Target=1
- Target=0

Source

Content Based Recommender System (RS)

Build experiences for millions of music lovers with playback, personalization, and much, much more.

Features

- Acousticness
- 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

SAS Enterprise Miner - Flux

Abstract
Introduction
Methods
Results
Conclusion

Model Comparison

![Receiver Operating Characteristic](image)
SAS® Enterprise Miner™ vs. Scikit-Learn – How do they recommend me good songs?

Raphael Lima

**Accuracy Score**

<table>
<thead>
<tr>
<th>Model</th>
<th>SAS® Enterprise Miner™</th>
<th>Scikit-Learn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>87.3%</td>
<td>83.9%</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>90.5%</td>
<td>87.3%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>90.5%</td>
<td>86.8%</td>
</tr>
<tr>
<td>Neural Net</td>
<td>90.2%</td>
<td>88.8%</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>89.3%</td>
<td>85.4%</td>
</tr>
<tr>
<td>SVM</td>
<td>89.3%</td>
<td>88.2%</td>
</tr>
<tr>
<td>D. Tree</td>
<td>88.1%</td>
<td>81.7%</td>
</tr>
</tbody>
</table>

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)

**Abstract**

Introduction

Methods

Results

Conclusion

**Logistic Regression**

**Ridge Penalty**

**Gradient Boosting**

**Random Forest**

**D. Tree**

**Neural Net**

**SVM**

**Naive Bayes**

**Embeded Variables Standardization**

**No Penalty**

**Depth: 2**

**Maximum number of trees**: 100

**T.Criterion: Chi Square**

**Hidden Layer Size**: 3

**Kernel**: linear

**Parenting Method**

**Standard Scaler Required**

**No Penalty**

**Depth: 3**

**Number of estimators**: 10

**T.Criterion: GINI**

**Hidden Layer Size**: 100

**Kernel**: rbf

**Priors = None**
Conclusion

In this case study, comparing classification methods, we concluded that with SAS Enterprise Miner, it was possible to build 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 assistance in this poster.

References

https://documentation.sas.com/?cdcid=emlearn&cdcVersion=1.0&docsetId=emex&docsetTarget=titlepage.htm&locale=pt-BR
Raphael Lima
Data Science & Analytics