

References and Links

General

- A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, O'Reilly Media, Inc. (2019).
- G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning*, Springer (2019).

Data Understanding

- [User-Agent request header information](#). MDN Web Docs.
- [Browser-Info page](#).
- [Wikipedia](#). Unix time.
- [HTTP Request Methods](#). W3 Schools.
- [HTTP Status Code Definitions](#).

Problem Background

- [Investopedia](#). Churn rate.
- [Heap Platform](#).
- P. Campbell [Ultimate Guide to Churn Rate](#), Profitwell Blog (Dec 2020).
- P. Campbell [Why a Customer Churn Model for Prediction is Challenging, But Worth it.](#), Profitwell Blog (May 2020).
- R. Law, [SaaS Companies Churn Rates](#), cobloom Blog (July 2020).

Machine Learning and Foundations

- P. Branco, L. Torgo, R.P. Ribeiro, *A survey of predictive modelling under imbalanced distributons*, arXivL 1505.01658v2 [cs.LG] (2015)
- J. Davis, M. Goadrich, *The relationship between precision-recall and ROC curves*, Proceedings of the 23rd International Conference on Machine Learning, Pittsburg, PA, 2006.
- J. Czakon, [F1-Score vs ROC-AUC vs Accuracy vs PR-AUC](#), [Neptune.ai](#) Blog.
- G. Azevedo, [Feature selection techniquesfor classification and Python tips for their application.](#), Medium (Aug. 2019).
- Kendall rank correlation coefficient. [Wikipedia page](#)
- [Python | Kendall Rank Correlation Coefficient](#), July 2021.
- [The ZeroR Classifier](#).
- A. Lee, [Choosing a Baseline Accuracy for a Classification Model](#), Medium (May 2021).
- J. Brownlee, [Stacking Ensemble Machine Learning with Python](#), Machine Learning Mastery (April 2021).
- J. Brownlee, [Essence of Stacking Ensembles for Machine Learning](#), Machine Learning Mastery (April 2021).

Spark

- A. Tandon, [Advanced Analytics with PySpark.](#), O'Reilly Media Inc. (April 2022) early release.
- D. Lee, T. Drabas, PySpark Cookbook, Packt Publishing (June 2018).
- J. Damji et al., Learning Spark, 2nd Edition, O'Reilley Media, Inc. (July 2020).
- [Guru 99: PySpark Tutorial for Beginners: Learn With Examples](#).
- [Spark Documentation.](#), Apache-Spark.
- Spark by Examples, [PySpark Tutorial For Beginners](#).

AWS EMR Cluster

- Brent Lemieux, [Getting Started with PySpark on AWS EMR](#), Medium (July 2019).
- Brent Lemieux, [Quick Setup Guide for your AWS Account](#).
- [AWS Big Data Blog](#). Install Python libraries on a running cluster with EMR Notebooks., Medium (March 2021).
- [Amazon EMR Guide](#). Installing and using kernels and libraries.
- A. Young, [How to set-up a cost-effective AWS EMR cluster and Jupyter notebooks for SparkSQL](#), Medium (Nov. 2020).