311-Service-Request Data Analysis using Apache Spark - SOEN 691 Project

Apoorv Semwal Hareesh Kavumkulath Loveshant Grewal

Abstract

*Recent advances in the field of Big Data Analytics and Machine Learning have introduced a plethora of open-source tools and technologies for both, academia and the massive data analyst community. In this project we try leveraging one such popular distributed data processing framework Apache Spark to analyse 311 - Service Request Data for the city of XXXXX. Being updated almost on a daily basis for the last XXXXX years, large volume of this dataset makes it a suitable candidate for analysis using a distributed processing framework like Spark. Making use of Spark Ecosystem libraries like Spark SQL and Spark ML, on this dataset, enables us to derive some interesting insights for better planning of city's resource. Identifying 3 primary goals for this project we first try answering a few statistical questions like "*most frequent complaints reported*", "*Average time to resolve the complaint/issue*". This involves making extensive use of Spark SQLs Dataframe API. Secondly we develop a predictive model for XXXXX, by evaluating performance of a set of selected supervised learning algorithms available in Spark ML. As part of our last goal we would be applying K-Means clustering over a selected set of features dividing the dataset into clusters and analysing them to identify any underlying service request patterns.*

1. Introduction

2. Section 2

1. Sample 1: Its some random text:

- Sample Item 1 ***Sample Italics*** in the dataset.
- Sample Item 1 ***Italics***,

2. Sample 2:

Sample Table	
Sample1	Sample_1, Sample_2, Sample_3_1
Sample1	Sample_1, Sample_2, Sample_3_1
Sample1	Sample_1, Sample_2, Sample_3_1

Sample Reference of a figure in text. Figure 1.

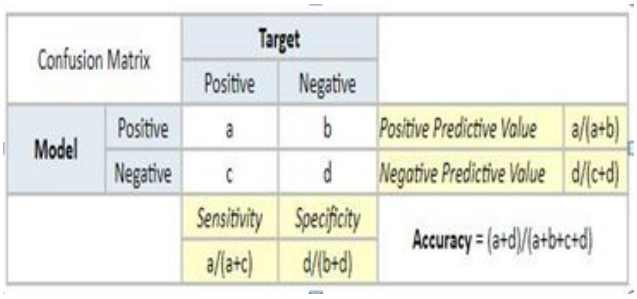

Figure 1. Sample Confusion Matrix.[1]

Multicol Table				
ID	Sample2	Sample3	Sample4	Sample5
1	74.5	64.5	63.5	66.0
2	78.41	81.61	81.81	82.95

$$\frac{\partial SampleEquation_{y_i}}{\partial X} \quad (1)$$

3. Conclusions

References

[1] TAVISH SRIVASTAVA. Analytics vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/, Aug 2019. 1