



COMP9313 Big Data Management

Project 2

Report

z5240221 Jatin Gupta

Table of Contents

TABLE OF CONTENTS.....	1
2 TASK 2	2
2.1.....	2
2.1.1	2
2.1.2	2
2.2.....	2

2 Task 2

2.1

2.1.1

For Test Data my stacking model F1 score is 0.7483312619309965

2.1.2

1. We can do the following
 - a. We can tune the hyper parameters in all the machine learning function given by pyspark. I tried working on Logistic regression class and when I tuned some parameters my f1 score improved the following:
 - i. F1 = 0.7495516561819556 at maxIter=5
 - ii. F1 = 0.7541015401421506 at maxIter=5, regParam=0.6
 - b. Upon changing the Naïvebayes class with the following hyper parameters I was able to get an even better score.
 - i. F1 = 0.7583873918357283 at smoothing=0.8 for all naïve bayes classes