**Hypotheses**

**Abhi**

**Hyp1 -> all numeric and all categorical features**

**Hyp2 -> core variables**

**Hyp3 → core + weather related features**

**Hyp4 → core + indicator/computed columns**

**Sarah**

**Hyp2 → Sarah’s core variables**

**Hyp3 → Sarah’s core + weather + percent variables**

**Hyp1 → Sarah’s core + weather + percent + mean variables**

**Hyp4 → Sarah’s core + custom (some weather + ind + mean + percentage)**

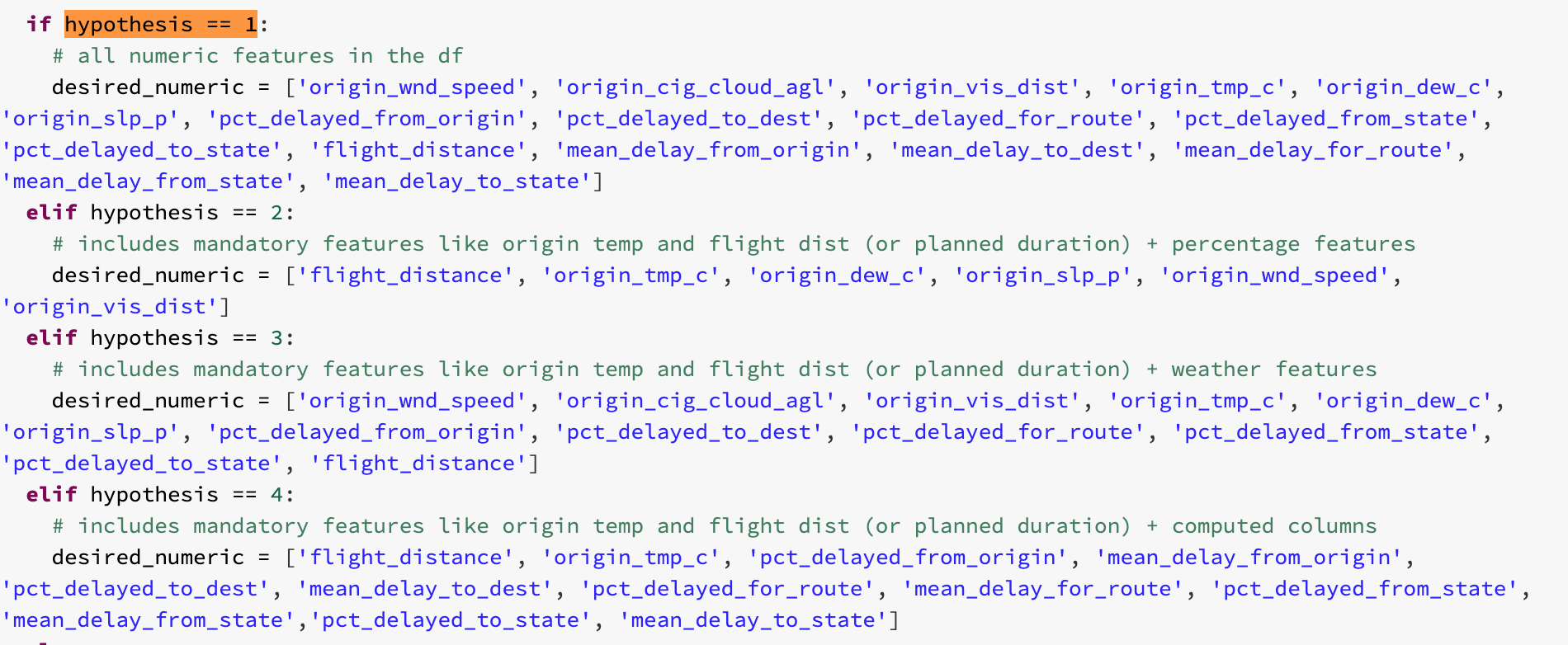
**Logit/SVM - made minor improvements in Hyp 4**

**RF -- made significant improvements in Hyp4**

**GBT - N/A**

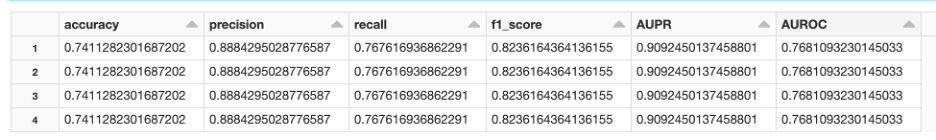
**Sarah Hypothesis**

****

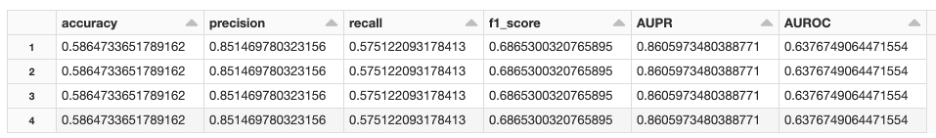
****

**Random forest**

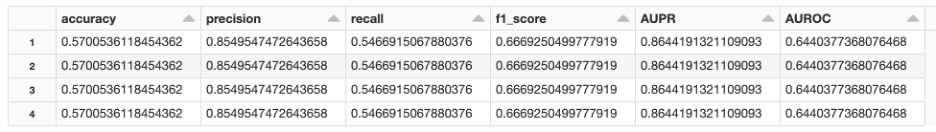
**Hypothesis 4**

****

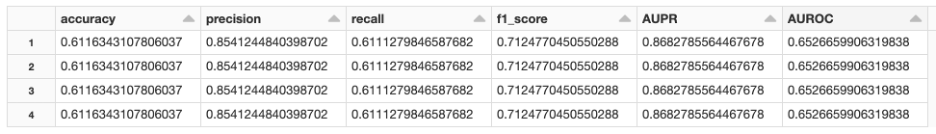
**Hypothesis 1**

****

**Hypothesis 3**

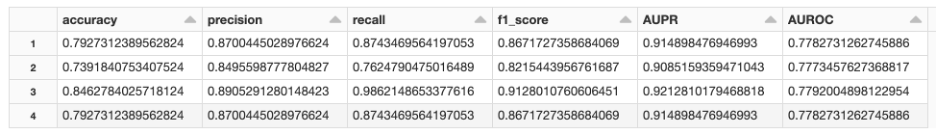
****

**Hypothesis 2**

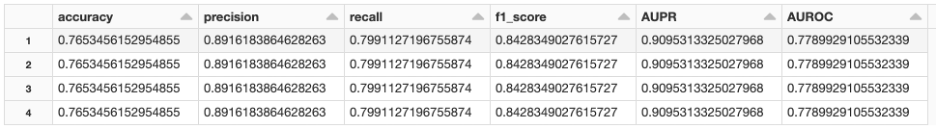
****

**SVM**

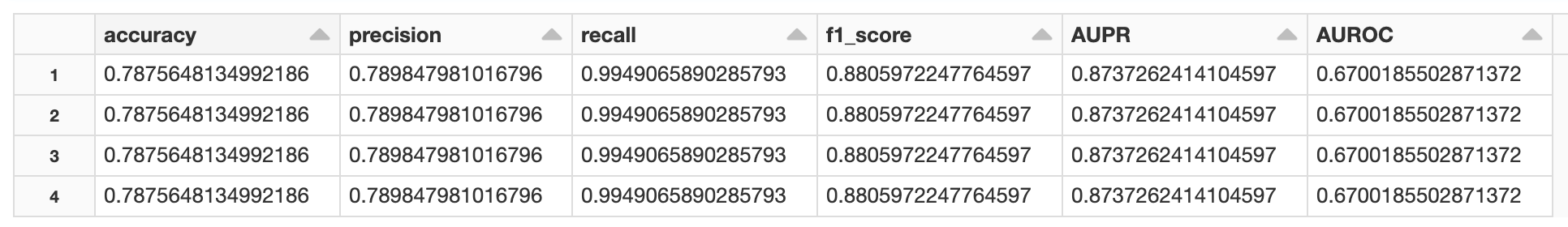
**Hypothesis 4 - 2 iterations**

****

**Hypothesis 4 - 1 iteration**

****

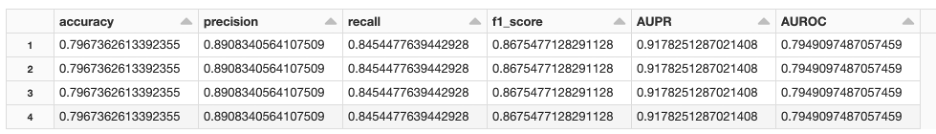
**Hypothesis 1 - iteration - got an error for this one for some reason**

**Hypothesis 3 - 1 iteration **

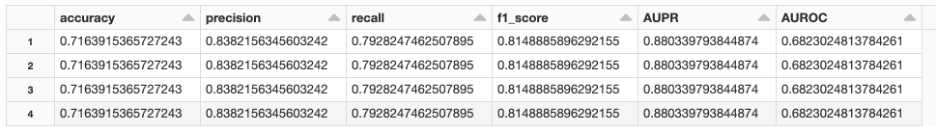
**Hypothesis 2 - 1 iteration**

**Logit**

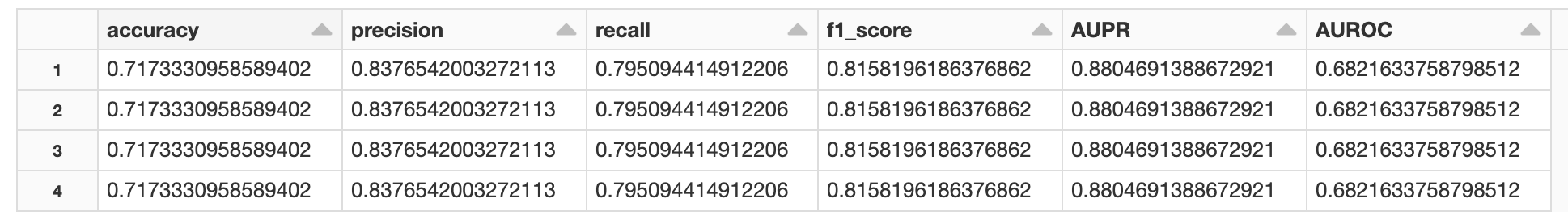
**Hypothesis 4**

****

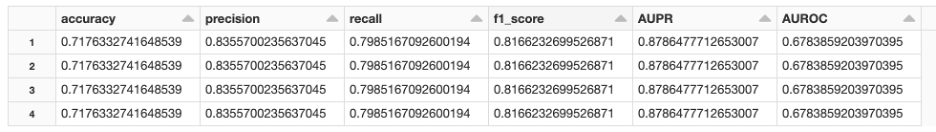
**Hypothesis 1**

****

**Hypothesis 3 - 1 iteration**

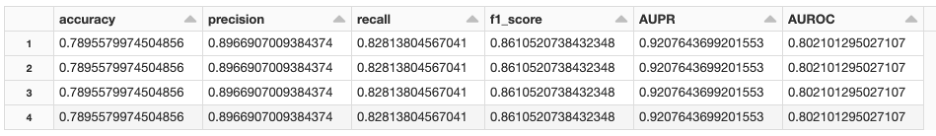
****

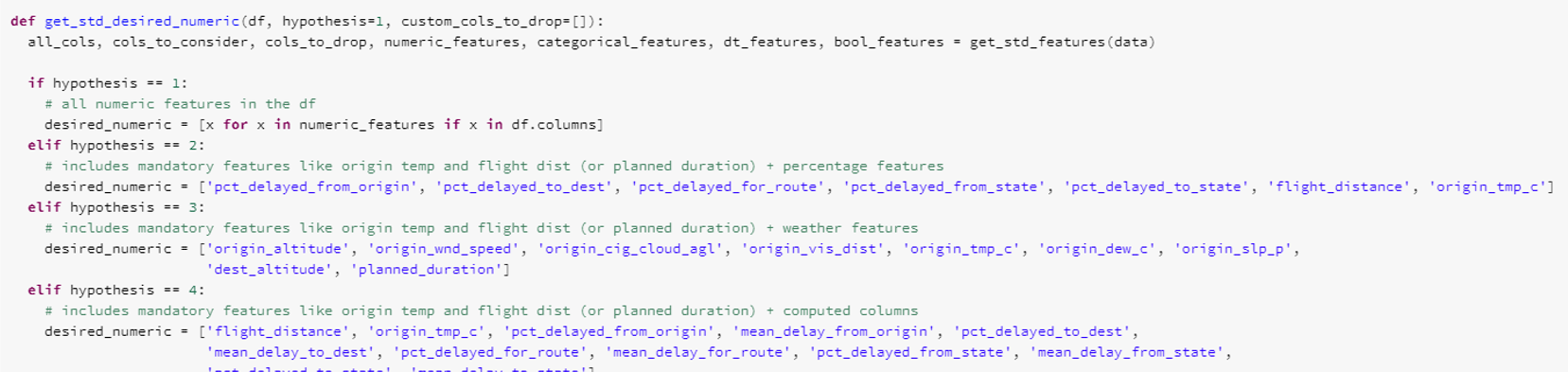
**Hypothesis 2**

****

**GBT**

**Hypothesis 4 - 1 iteration**

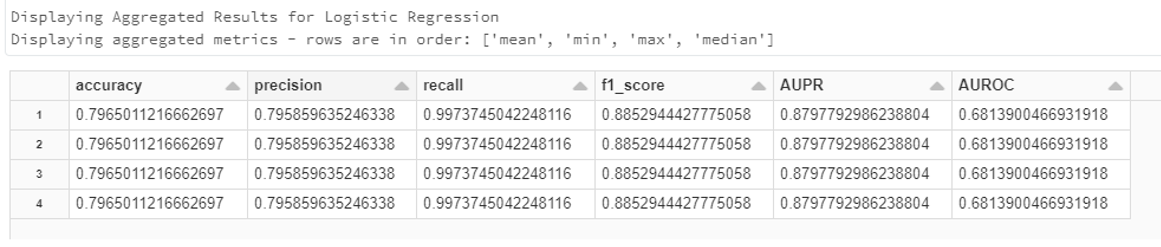
****

**Abhi Hypothesis**

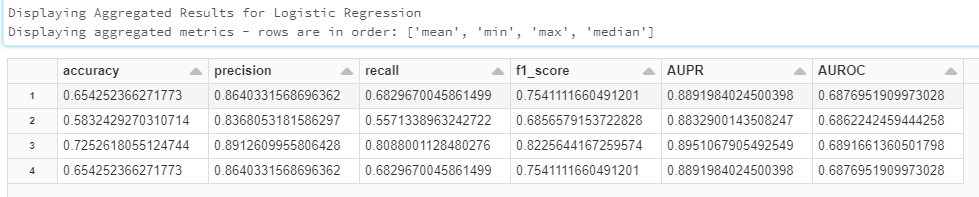


**Logistic Regression**

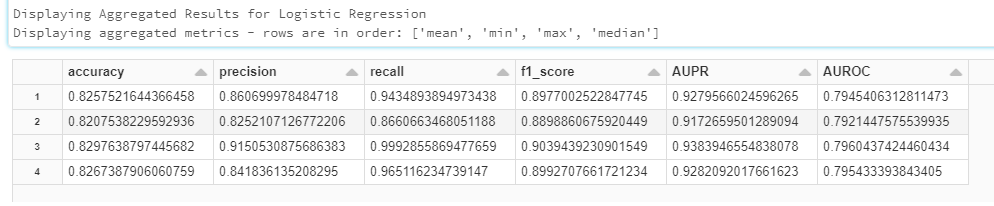
*Abhi: Hypothesis 2 – 1 iteration – Default Params - Runtime: 5:99 min*



*Abhi: Hypothesis 3 - 2 iterations - Default Params - Runtime: 10.42 min*

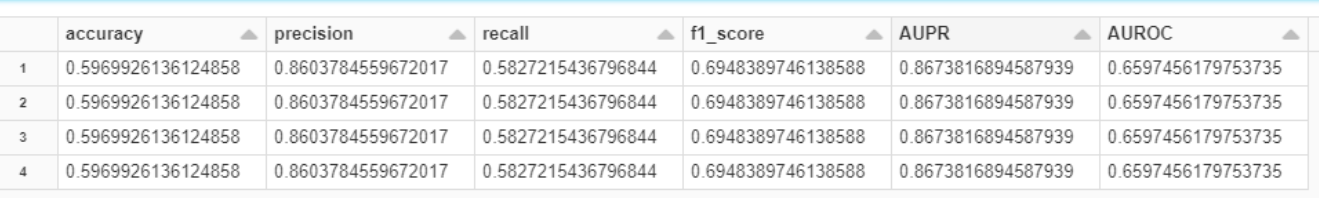


*Abhi: Hypothesis 4 - 3 iterations - Default Params - Runtime: 26 min*

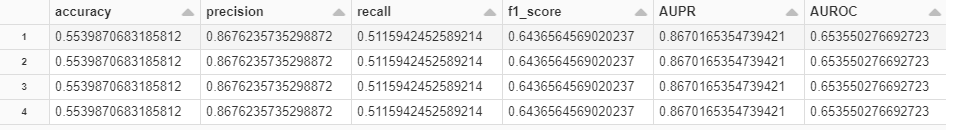


**Random Forest**

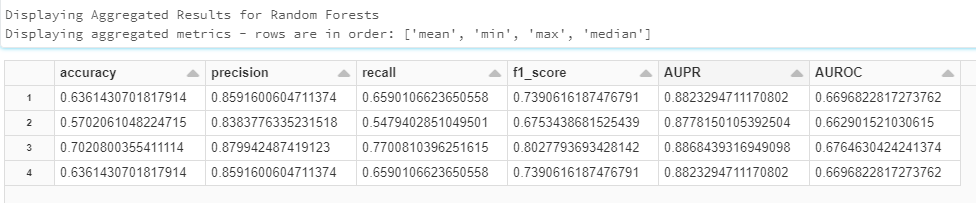
*Abhi: Hypothesis 2 – 1 iteration – Default Params - Runtime: 3:34 min*



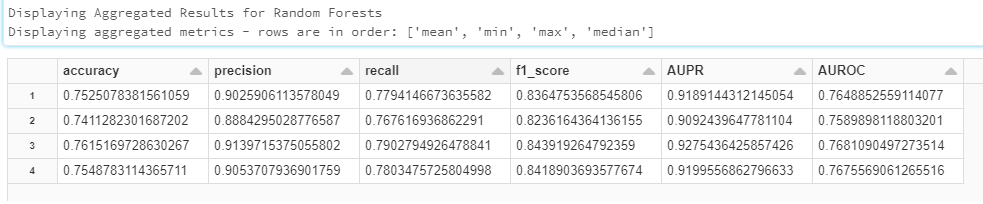
*Abhi: Hypothesis 2 - 1 iteration - Default Params - Runtime: 3 min* ***(WITH QUANTIZER)***

******

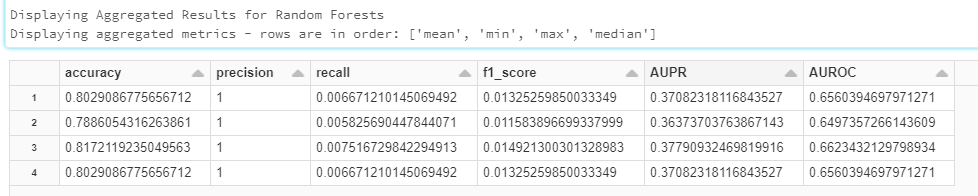
*Abhi: Hypothesis 3 – 2 iteration – Default Params - Runtime: 6:05 min*



*Abhi: Hypothesis 4 – 3 iteration – Default Params - Runtime: 35:55 min*

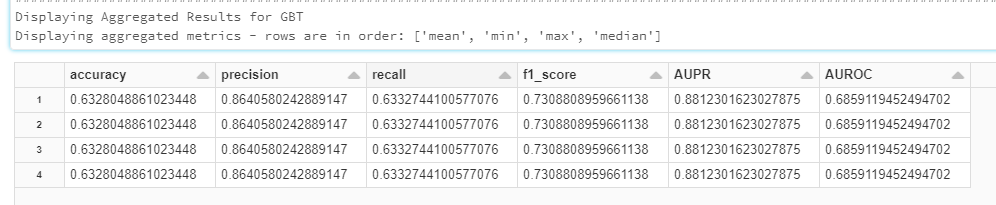


Abhi: Hypothesis 3 - 2 iteration - Default Params - Runtime: 33.77 mins **(WITH REBALANCE = FALSE - NOTICE HOW RECALL GOES TO SHIT)**

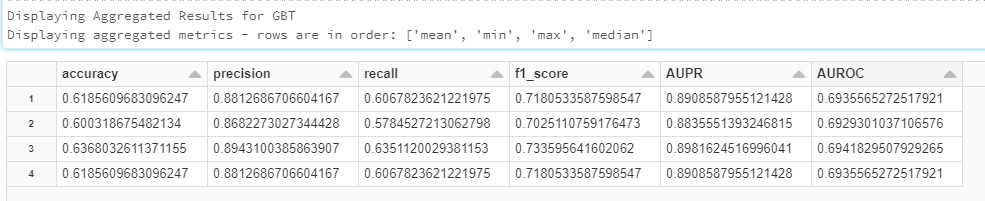


**Gradient Boosted Trees**

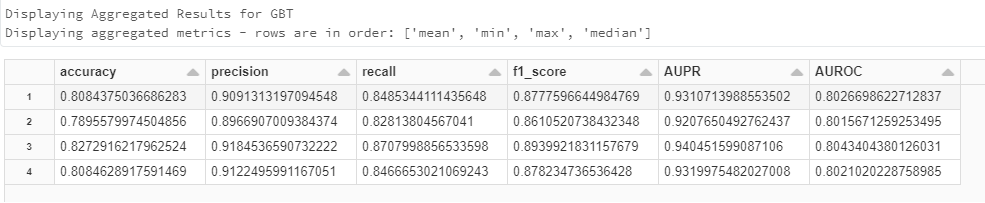
*Abhi: Hypothesis 2 – 1 iteration – Default Params - Runtime: 20:18 min*



*Abhi: Hypothesis 3 - 2 iteration - Default Params - Runtime: 1:10* ***hour***

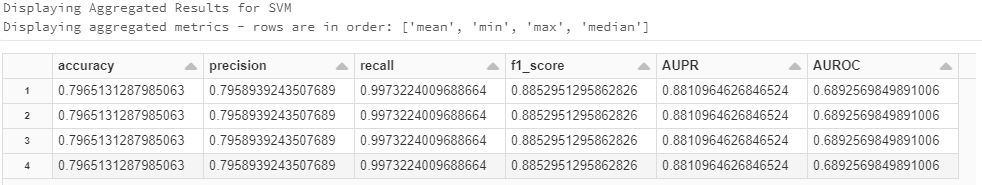
******

*Abhi: Hypothesis 4 - 3 iteration - Default Params - Runtime: 1:30* ***hour***



**Support Vector Machine**

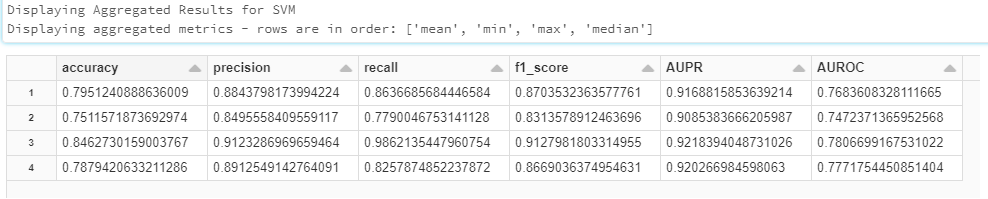
*Abhi: Hypothesis 2 – 1 iteration – Default Params - Runtime: 5:03 min*



*Abhi: Hypothesis 3 – 2 iteration – Default Params - Runtime: xx min*

Errored out - for some reason - precision/recall/f1 were 0 - need to investigate if we have a bad feature in there

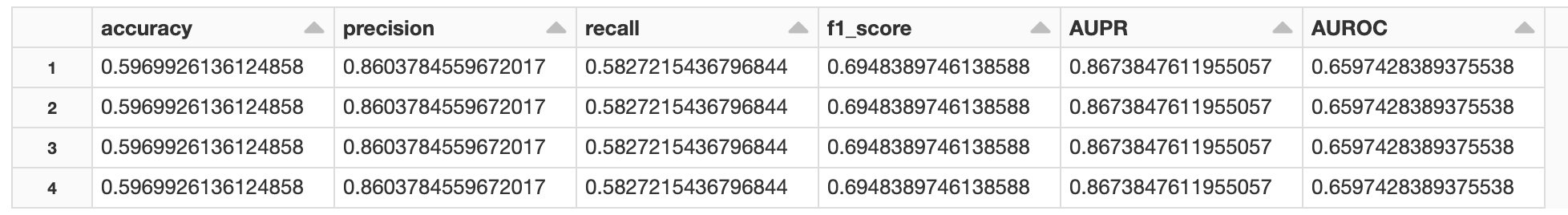
*Abhi: Hypothesis 4 - 3 iteration - Default Params - Runtime: 1:30* ***hour***



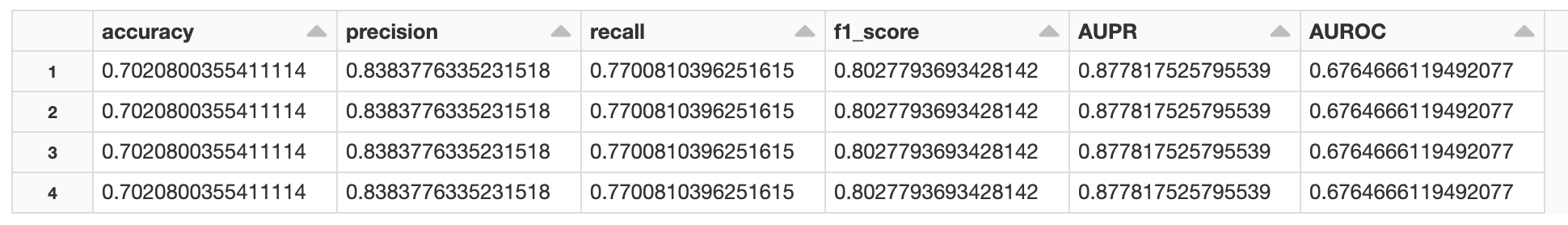
Sarah

**Random Forest**

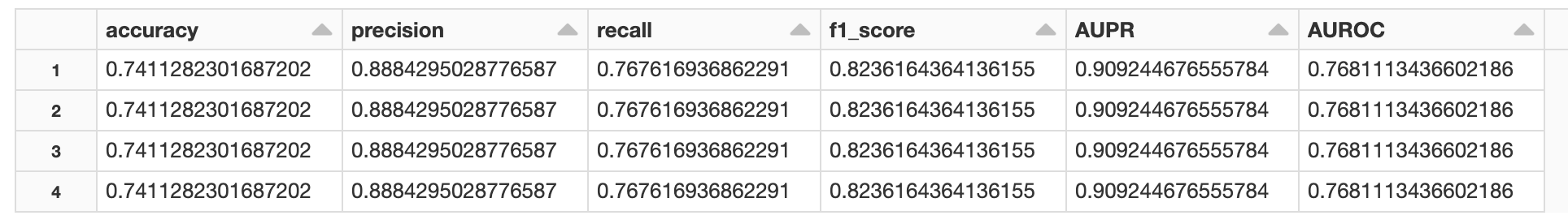
Hypothesis 2 - 1 iteration - Max Bins=40, Num Trees=15, Min Info Gain=0.2, Min Instances Per Node=DEFAULT



Hypothesis 3 - 1 iteration - Max Bins=40, Num Trees=15, Min Info Gain=0.2, Min Instances Per Node=DEFAULT

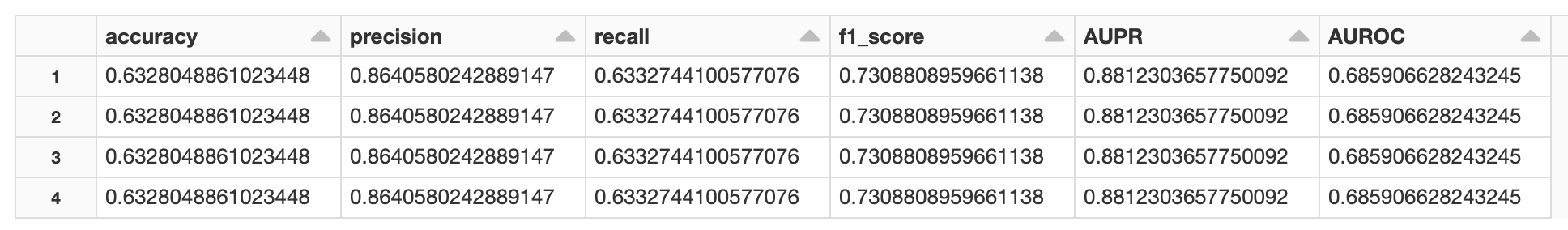


Hypothesis 4 - 1 iteration - Max Bins=40, Num Trees=15, Min Info Gain=0.2, Min Instances Per Node=DEFAULT

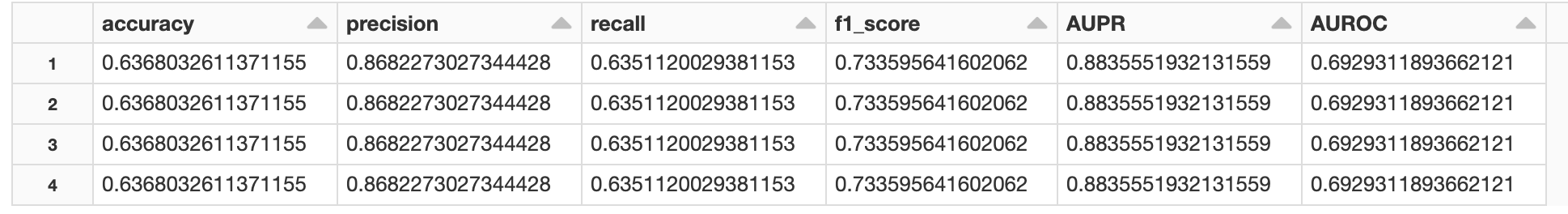


**Gradient Boost**

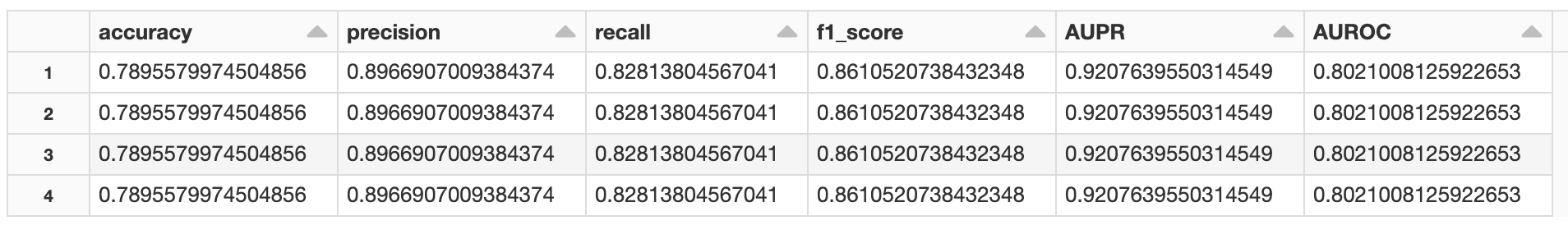
Hypothesis 2 - 1 iteration - Max Bins=40, Max Depth =15, Min Instance Per Node = DEFAULT, minInfoGain=0.2, MaxIter=, step size = 0.1



Hypothesis 3 - 1 iteration - Max Bins=40, Max Depth =15, Min Instance Per Node = DEFAULT, minInfoGain=0.2, MaxIter=, step size = 0.1

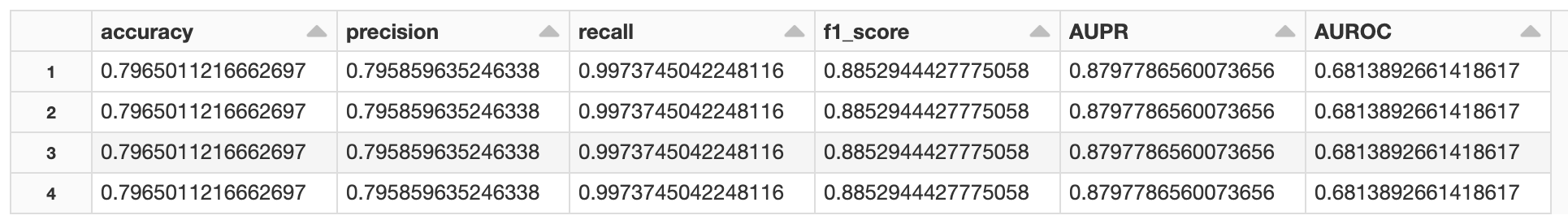


Hypothesis 4 - 1 iteration - Max Bins=40, Max Depth =15, Min Instance Per Node = DEFAULT, minInfoGain=0.2, MaxIter=, step size = 0.1

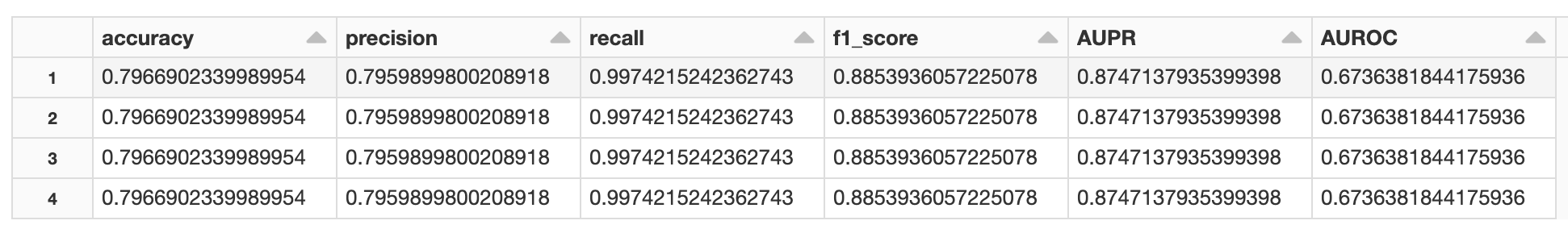


**Logistic Regression**

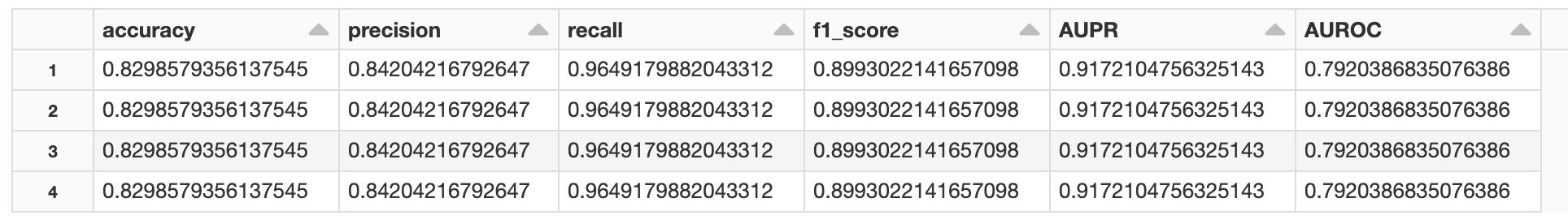
Hypothesis 2 - 1 iteration - threshold (>0.5 assigns to class 1= 0.6, maxIter=15, regParam=0.1



Hypothesis 3 - 1 iteration - threshold (>0.5 assigns to class 1= 0.6, maxIter=15, regParam=0.1



Hypothesis 4 - 1 iteration - threshold (>0.5 assigns to class 1= 0.6, maxIter=15, regParam=0.1



**SVM**

Hypothesis 2 - 1 iteration