Parameters:

Kmer\_size:

Bloom\_tree threshold

Size of bloom\_filter

Number of hash functions

FPR when using formulas

Results:

Accuracy of bad\_data identification

Rate of good\_data identification

Time

Space (sys.getsizeof)

Confusion Matrix

Things to consider:

Avg number of matches

How filled in the bloom filter is

1 hash function

m = 1000

kmer\_size = 30

threshold = 0.3

Accuracy: 1.0

AvgMatches: 5.0

Fpr = 0.1

BFsize: 120035574

Kmer\_size = 30

Threshold = 0.3

Hash Count: 3

Accuracy: 1.0

AvgMatches: 1.048

Kmer\_size = 30

Fpr = 0.1

Threshold = 0.3

Accuracy: 1.0

AvgMatches: 1.048

avgMouseMatches: 0.0

BF size: 120035574

Hash Count: 3

Bloom Tree size: 56 bytes

Kmer\_size = 5

Fpr = 0.1

Threshold = 0.3

Accuracy: 1.0

AvgMatches: 5.0

avgMouseMatches: 5.0

BF size: 120057140

Hash Count: 3

Bloom Tree size: 56

Kmer\_length: 30

FPR: 0.1

Query Threshold: 0.3

Accuracy: 1.0

AvgMatches: 1.048

avgMouseMatches: 0.0

Bloom Filter size: 120035574

Hash Count: 3

Bloom Tree size: 56

Construction Time: 623.1445491313934

Query Time: 263.23813796043396