One-class SVM using Binary, Frequency, TF-IDF and Hadamard Representations, with Different Vector Dimensions, and Measures (F1-measure, Recall, Precision)

Figure 1 – Table of One-class SVM using Vector Size of 15 on Various Representations

Keywords Vector Size $(m) = 15$						
Kernel	*Linear*			*Radial* (RBF)		
Representation	F_1	R	P	F_1	R	P
Binary	0.638	0.744	0.558	0.773	0.902	0.676
Frequency	0.550	0.641	0.481	0.737	0.860	0.645
TF-IDF	0.673	0.785	0.589	0.703	0.820	0.615
Hadamard	0.564	0.658	0.493	0.725	0.846	0.634
∗ Value of each cell is the average of 100 samples of identical nature						

Figure 2 – Table of One-class SVM using Vector Size of 30 on Various Representations

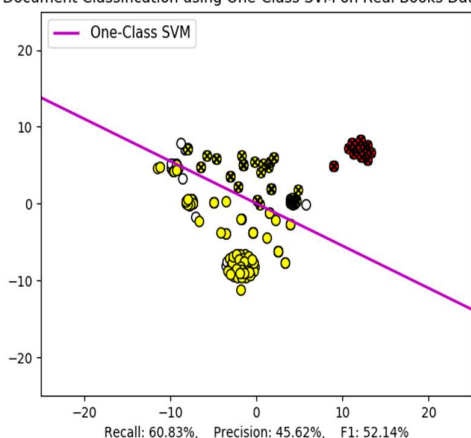| *Keywords Vector Size* $(m) = 30$						
Kernel	*Linear*			*Radial* (RBF)		
Representation	F_1	R	P	F_1	R	P
Binary	0.505	0.589	0.442	0.785	0.916	0.687
Frequency	0.580	0.677	0.507	0.628	0.732	0.549
TF-IDF	0.673	0.785	0.589	0.725	0.845	0.634
Hadamard	0.554	0.646	0.484	0.717	0.837	0.628
∗ Value of each cell is the average of 100 samples of identical nature						

Figure 3 – Binary Representation in Linear and Radial (RBF) Kernels

* WHITE – Training Data   * YELLOW – POSITIVE DATA   * RED – NEGATIVE DATA

Figure 4 – Frequency Representation in Linear and Radial (RBF) Kernels

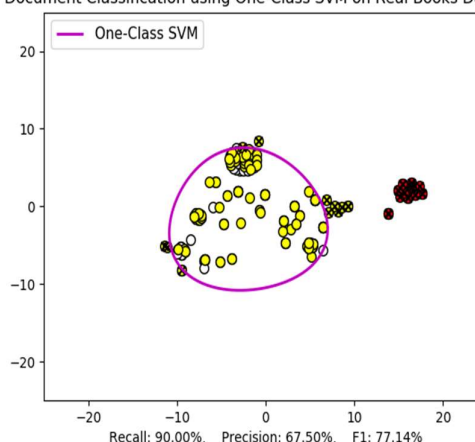* WHITE – Training Data   * YELLOW – POSITIVE DATA   * RED – NEGATIVE DATA

Document Classification using One-Class SVM on Real Books Data Set   Document Classification using One-Class SVM on Real Books Data Set
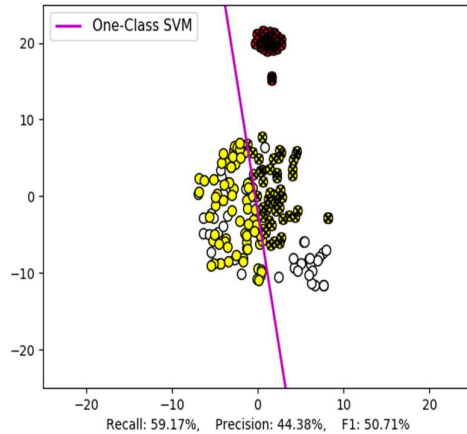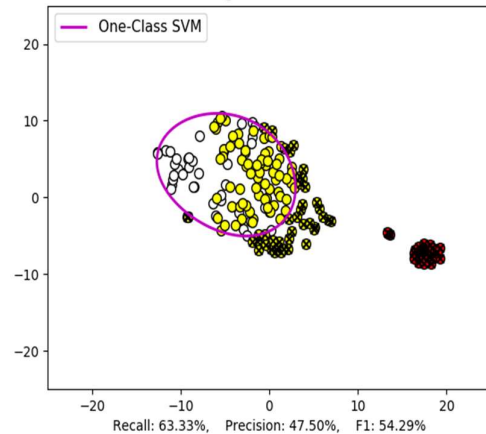
Recall: 59.17%,   Precision: 44.38%,   F1: 50.71%   Recall: 63.33%,   Precision: 47.50%,   F1: 54.29%

Figure 5 – TF-IDF Representation in Linear and Radial (RBF) Kernels

* WHITE – Training Data   * YELLOW – POSITIVE DATA   * RED – NEGATIVE DATA

Document Classification using One-Class SVM on Real Books Data Set   Document Classification using One-Class SVM on Real Books Data Set
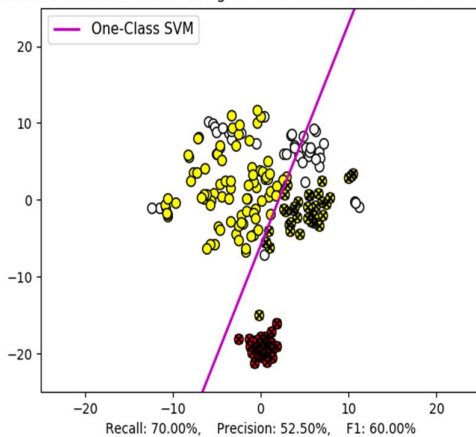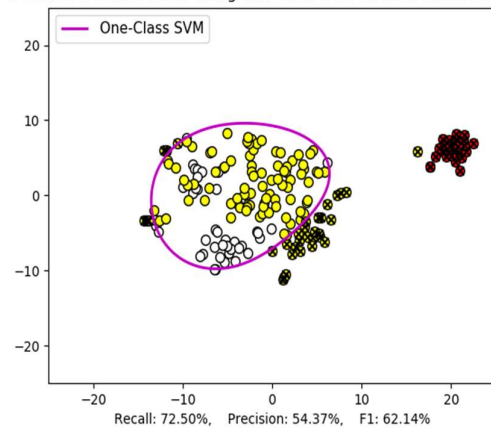
Recall: 70.00%,   Precision: 52.50%,   F1: 60.00%   Recall: 72.50%,   Precision: 54.37%,   F1: 62.14%

Figure 6 – Hadamard Representation in Linear and Radial (RBF) Kernels

* WHITE – Training Data   * YELLOW – POSITIVE DATA   * RED – NEGATIVE DATA

Document Classification using One-Class SVM on Real Books Data Set   Document Classification using One-Class SVM on Real Books Data Set
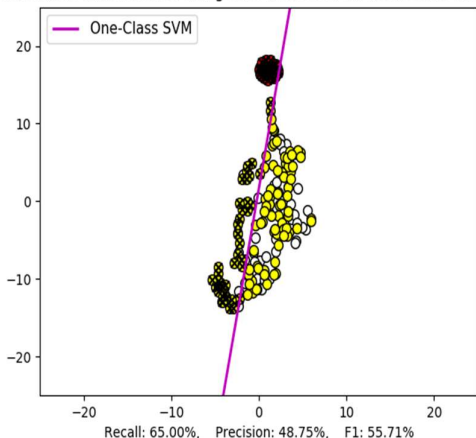
Recall: 65.00%,   Precision: 48.75%,   F1: 55.71%   Recall: 85.83%,   Precision: 64.38%,   F1: 73.57%