

Table 1: F<sub>1</sub> Scores for the *uncertain* class of the Reports dataset according to all six machine learning algorithms with tf weighting. Best results are presented in bold.

Dictionary	LR	SMO	JRIP	J48	RF	CNN
<i>Unc</i>	0.53	0.53	<b>0.55</b>	0.42	0.58	0.53
<i>UncGen</i>	0.55	<b>0.60</b>	0.55	0.47	0.43	0.52
<i>UncGen<sub>exp</sub></i>	<b>0.59</b>	0.53	0.58	0.42	0.58	0.55
<i>UncSpec</i>	0.42	0.55	<b>0.58</b>	0.49	0.26	0.47
<i>UncSpec<sub>exp</sub></i>	<b>0.59</b>	0.53	0.58	0.42	0.58	0.55
<i>UncGen+UncSpec</i>	0.42	0.55	<b>0.58</b>	0.49	0.26	0.47
<i>UncGen<sub>exp</sub>+UncSpec<sub>exp</sub></i>	<b>0.59</b>	0.53	0.58	0.42	0.58	0.55
Average	0.53	0.55	<b>0.57</b>	0.45	0.47	0.52

Table 2: F<sub>1</sub> Scores for the *uncertain* class of the Reports dataset according to all six machine learning algorithms with tf-idf weighting. Best results are presented in bold.

Dictionary	LR	SMO	JRIP	J48	RF	CNN
<i>Unc</i>	<b>0.59</b>	0.53	0.58	0.42	0.58	0.55
<i>UncGen</i>	0.59	0.53	<b>0.63</b>	0.42	0.58	0.56
<i>UncGen<sub>exp</sub></i>	0.42	0.55	<b>0.58</b>	0.49	0.26	0.47
<i>UncSpec</i>	0.55	<b>0.60</b>	0.55	0.47	0.43	0.53
<i>UncSpec<sub>exp</sub></i>	<b>0.59</b>	0.53	0.56	0.42	0.58	0.55
<i>UncGen+UncSpec</i>	0.49	0.54	<b>0.58</b>	0.52	0.35	0.50
<i>UncGen<sub>exp</sub>+UncSpec<sub>exp</sub></i>	0.53	0.53	0.55	0.42	<b>0.58</b>	0.53
Average	0.54	0.55	<b>0.58</b>	0.45	0.48	0.53