

MALWARE PREDICTION PROJECT RESULT

Model

- CatBoost
- Random Forest

Features

- All settings
- All features except settings
- All features

Table 1. Ranking of ML models by AUC

Model	Features	AUC in testing set
CatBoost	All features	0.70
CatBoost	All features except settings	0.68
CatBoost	All settings	0.66
Random Forest	All features	0.61
Random Forest	All features except settings	0.61
Random Forest	All settings	0.60

Figure 1. Normalized confusion matrix

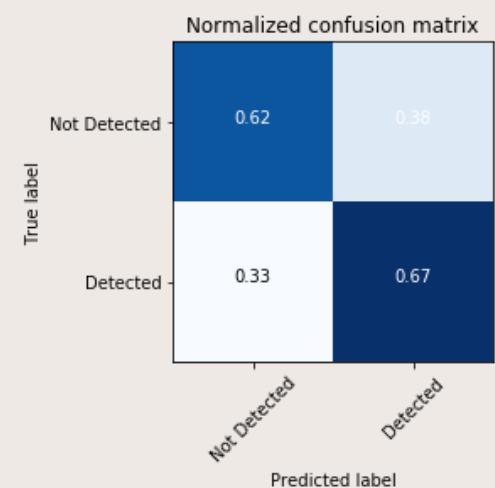


Figure 2. ROC curve

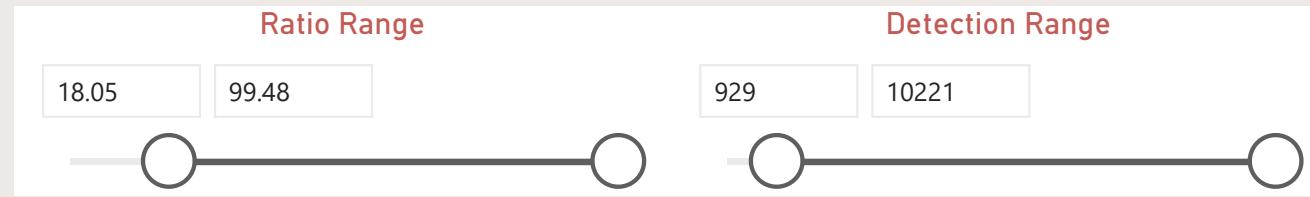
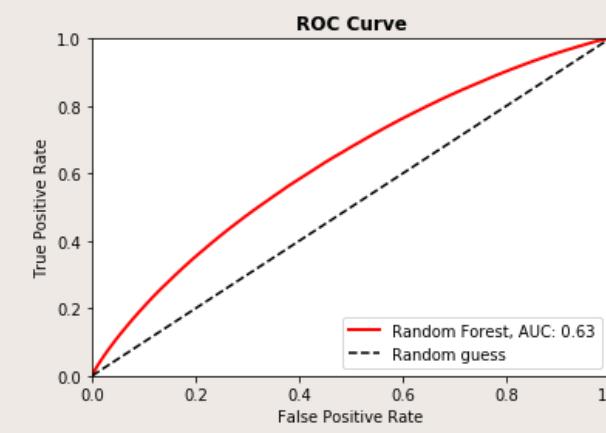


Table 2. Features and values ordered by detection count

Setting1	Value1	Setting2	Value2	Total	Detection	Ratio
Census_OEMModelIdentifier	40801	Census_OEMNameIdentifier	2144	3293	2068	62.00
Census_OSBuildRevision	709422	Census_OSVersion	708698	3552	2468	69.00
Census_OSBuildRevision	709425	Census_OSVersion	708701	3274	2014	61.00
Census_OSInstallLanguageIdentifier	709670	Census_OSUILocaleIdentifier	709707	17068	10221	59.00
OsBuildLab	608	Census_OSBuildRevision	709425	2802	1759	62.00
OsBuildLab	608	Census_OSVersion	708701	2802	1759	62.00
RtpStateBitfield	11	SmartScreen	1452	3840	3395	88.00
SmartScreen	1452	Census_OEMModelIdentifier	45506	1085	945	87.00
SmartScreen	1452	Census_OSInstallLanguageIdentifier	709670	4207	3696	87.00
SmartScreen	1452	Census_OSUILocaleIdentifier	709707	4293	3771	87.00
SmartScreen	1452	LocaleEnglishNameIdentifier	164	4034	3154	78.00
SmartScreen	1452	LocaleEnglishNameIdentifier	213	4420	3547	80.00
SmartScreen	1452	RtpStateBitfield	11	15360	13580	88.00

Figure 3. Detection by Setting1 and Value1

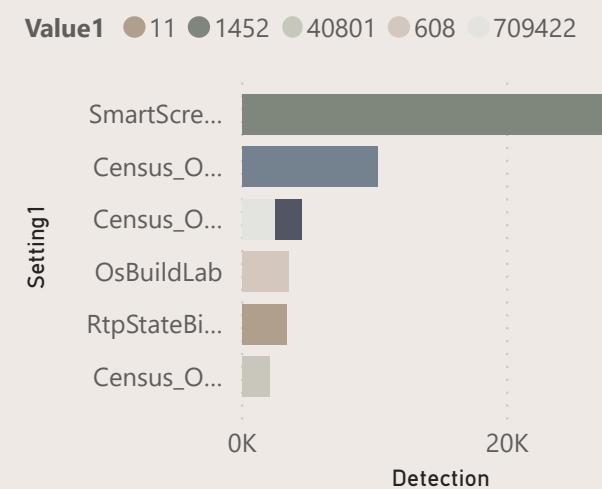


Figure 4. Detection by Setting2 and Value2

