УДК 629.78

ВЛИЯНИЕ ХАРАКТЕРИСТИК ПК НА ПРЕДСКАЗАНИЕ ВРЕДОНОСНОСТИ ПО

© Баканов Д.С., Лобанков А.А.

Самарский национальный исследовательский университет имени академика С.П. Королева, г. Самара, Российская Федерация

e-mail: dima.bakanov.1999@mail.ru

Вредоносное программное обеспечение (вредоносное ПО) – это собирательный термин, обозначающий вредоносную программу или часть кода [1]. Вредоносное ПО стремится проникнуть в систему и нанести урон, перехватить управление или полностью захватить контроль над некоторыми процессами или вывести из строя компьютеры, компьютерные системы, сети, а также мобильные устройства. Определенные устройства в зависимости от их характеристик (версия ОС, состояние антивируса и т. д.) могут быть более подвержены заражению вредоносным ПО.

В этой работе анализируются данные, собранные компанией Microsoft при помощи Защитника Windows (Windows Defender) [2]. Данные представлены в виде таблицы, в строках которой содержится информация о конкретном ПК и результат о выявлении вредоносного ПО (рис. 1) .

	MachineIdentifier	ProductName	EngineVersion	AppVersion	AvSigVersion	IsBeta	RtpStateBitfield	IsSxsPassiveMode	I
0	0000028988387b115f69f31a3bf04f09	win8defender	1.1.15100.1	4.18.1807.18075	1.273.1735.0	0	7.0	0	
1	000007535c3f730efa9ea0b7ef1bd645	win8defender	1.1.14600.4	4.13.17134.1	1.263.48.0	0	7.0	0	
2	000007905a28d863f6d0d597892cd692	win8defender	1.1.15100.1	4.18.1807.18075	1.273.1341.0	0	7.0	0	
3	00000b11598a75ea8ba1beea8459149f	win8defender	1.1.15100.1	4.18.1807.18075	1.273.1527.0	0	7.0	0	
4	000014a5f00daa18e76b81417eeb99fc	win8defender	1.1.15100.1	4.18.1807.18075	1.273.1379.0	0	7.0	0	

Рис. 1. Частичный внешний вид данных от компании Microsoft

В названии столбцов представлены признаки, включающие характеристики ПК (например, версия ОС) и характеристики антивируса (например, название антивируса). Также есть один признак, который показывает зараженность ПК (HasDetections). Данный признак зависим от остальных признаков и называется исходом. Исход принимает два значения: 1 – выявлено вредоносное ПО, 0 – не выявлено. Количество записей о ПК примерно пополам разделилось в зависимости от значений исхода, т. е. в данных нет дисбаланса между строками с HasDetections, равным 1 и 0. Среди данных были выявлены признаки, которые сильно не влияют на исход. К таким признакам были отнесены те, у которых доля пропущенных или часто встречаемых значений больше 0,9.

Для выявления лучшего предиктора исхода использовалась логистическая регрессия:

$$\ln\left(Шансы\left(Y=1\right)\right)=\beta_0+\beta_1 x_1,$$

где β_i – коэффициенты регрессии, x_1 – значения предикторов, *Шансы* – вероятность события Y, деленная на то, что данное событие не произойдет [3].

Предварительно признаки были разделены на категориальные, представленные в виде конечного количества строчных значений, и числовые. Среди категориальных признаков были выявлены следующие хорошие предикторы: версия (EngineVersion) защитника Windows и состояние функции SmartScreen (рис. 2).

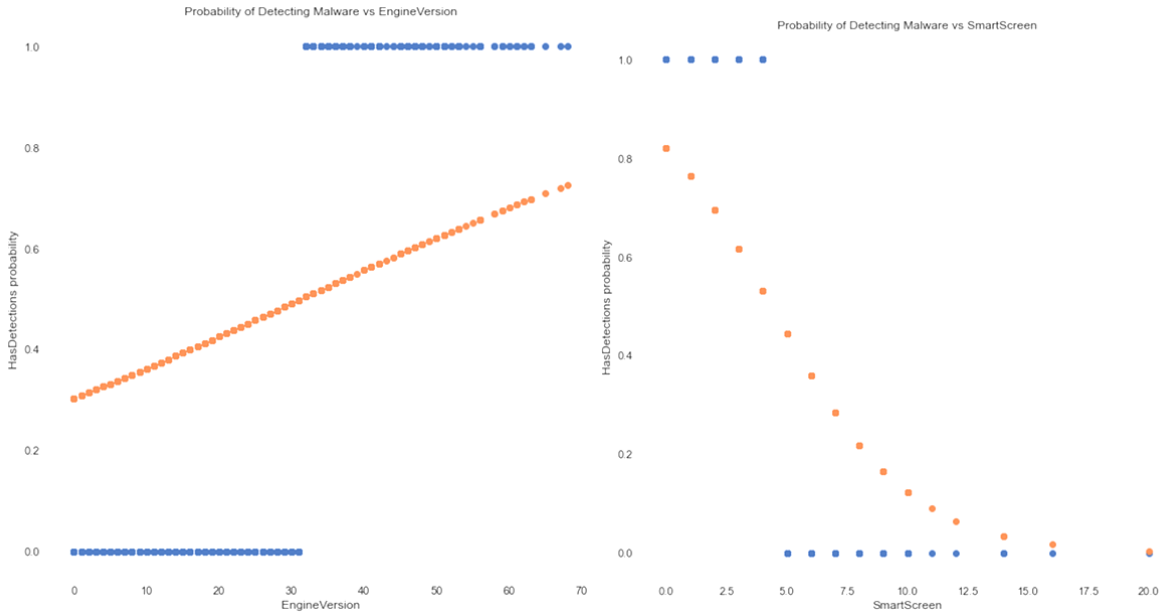

Рис. 2. Лучшие результаты логистической регрессии среди категориальных признаков

Среди числовых – количество логических ядер (Census_ProcessorCoreCount) и признак (AVProductsInstalled), указывающий, установлен ли антивирус (рис. 3).

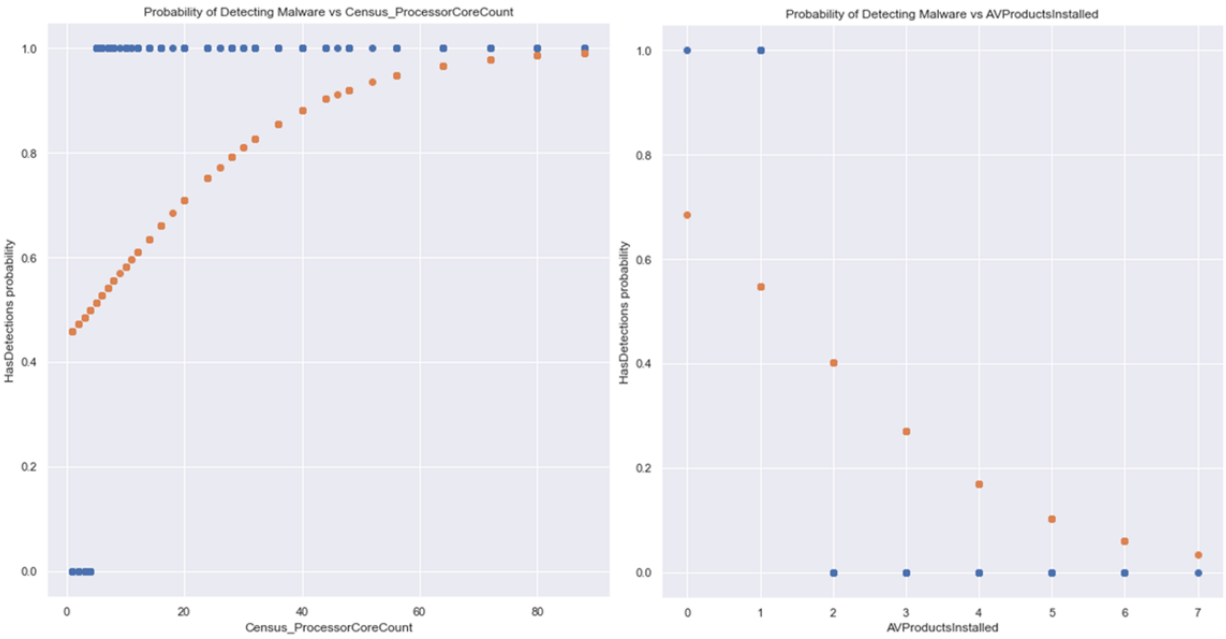

Рис. 3. Лучшие результаты логистической регрессии среди числовых признаков

Библиографический список

1. Вредоносное ПО // Malwarebytes. URL: https://ru.malwarebytes.com/malware (дата обращения: 25.10.2020).

2. Malware Prediction // Kaggle. URL: https://www.kaggle.com/c/microsoft-malware-prediction (дата обращения: 20.10.2020).

3. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science: пер. с англ. СПб.: БХВ-Петербург, 2020. 304 с.: ил.