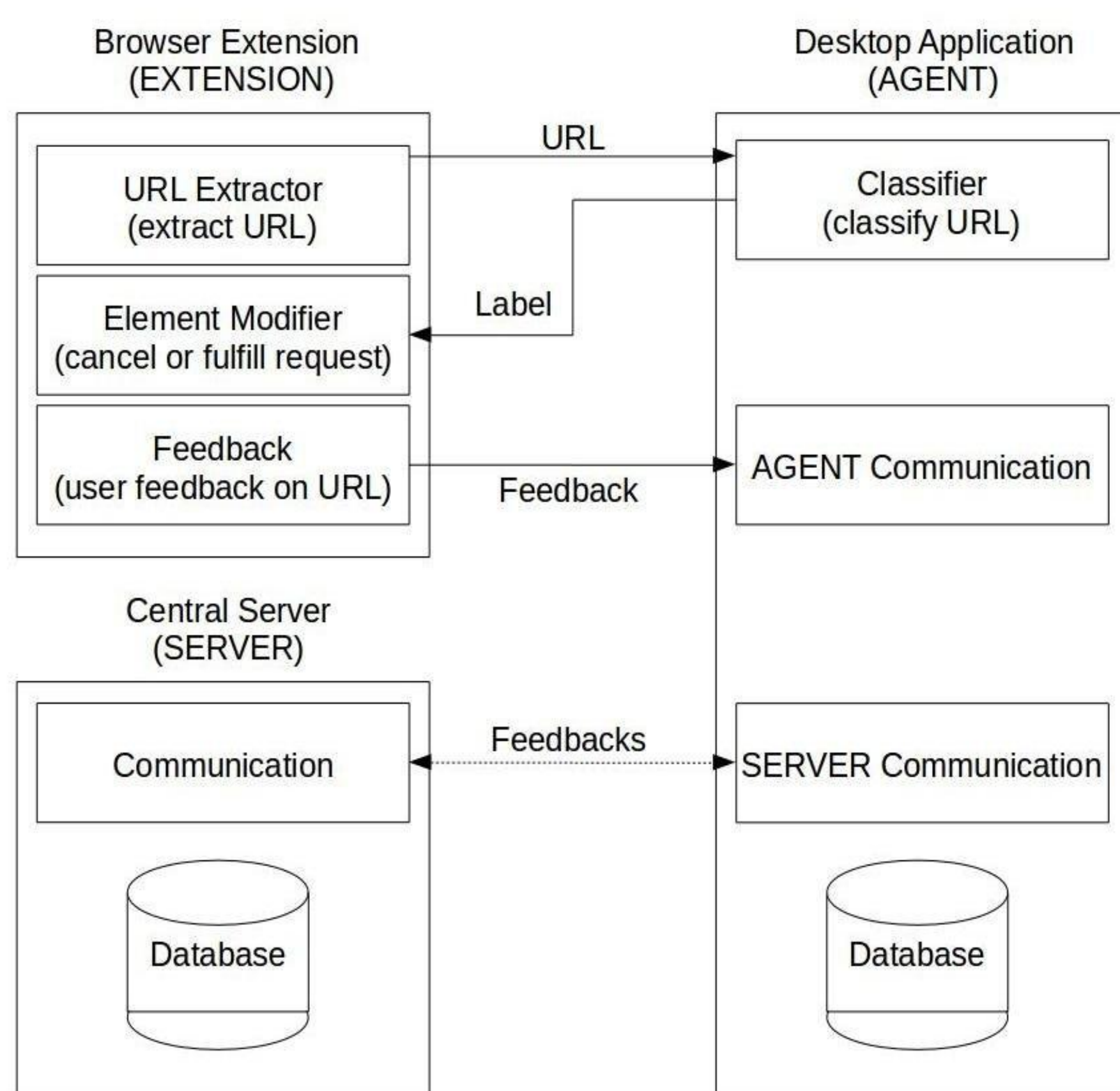


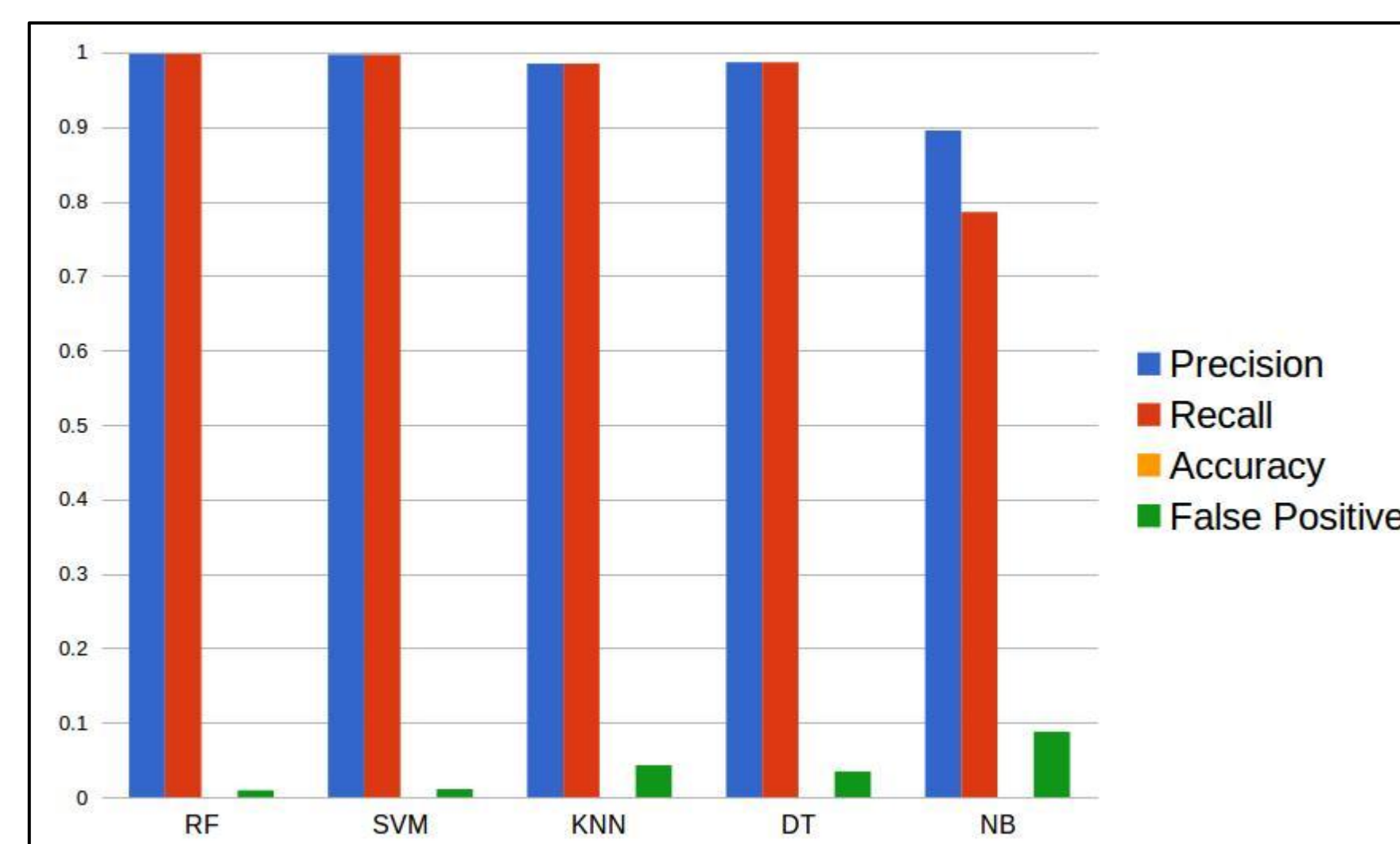
ABSTRACT

Online advertisements (ads) have taken over web and most websites contain some sort of ads. While ads produce revenue for the server maintainer or to businesses, they have become intrusive and dangerous as ever. Ads use more bandwidth, show inappropriate ads, and spread malware such as adware and ransomware. Although there are many products to block ads, also known as ad blockers, most depend on a filter list that must be managed manually. When malicious advertisers can produce millions of new URLs within minutes, this is not the most effective method against ads. We propose our own ad blocker, CIC-AB, which uses machine learning techniques to detect new and unknown ads without needing to update the filter list. CIC-AB showed highest accuracy and precision of 99.7% and false positive rate of 0.9%.

DEVELOPMENT



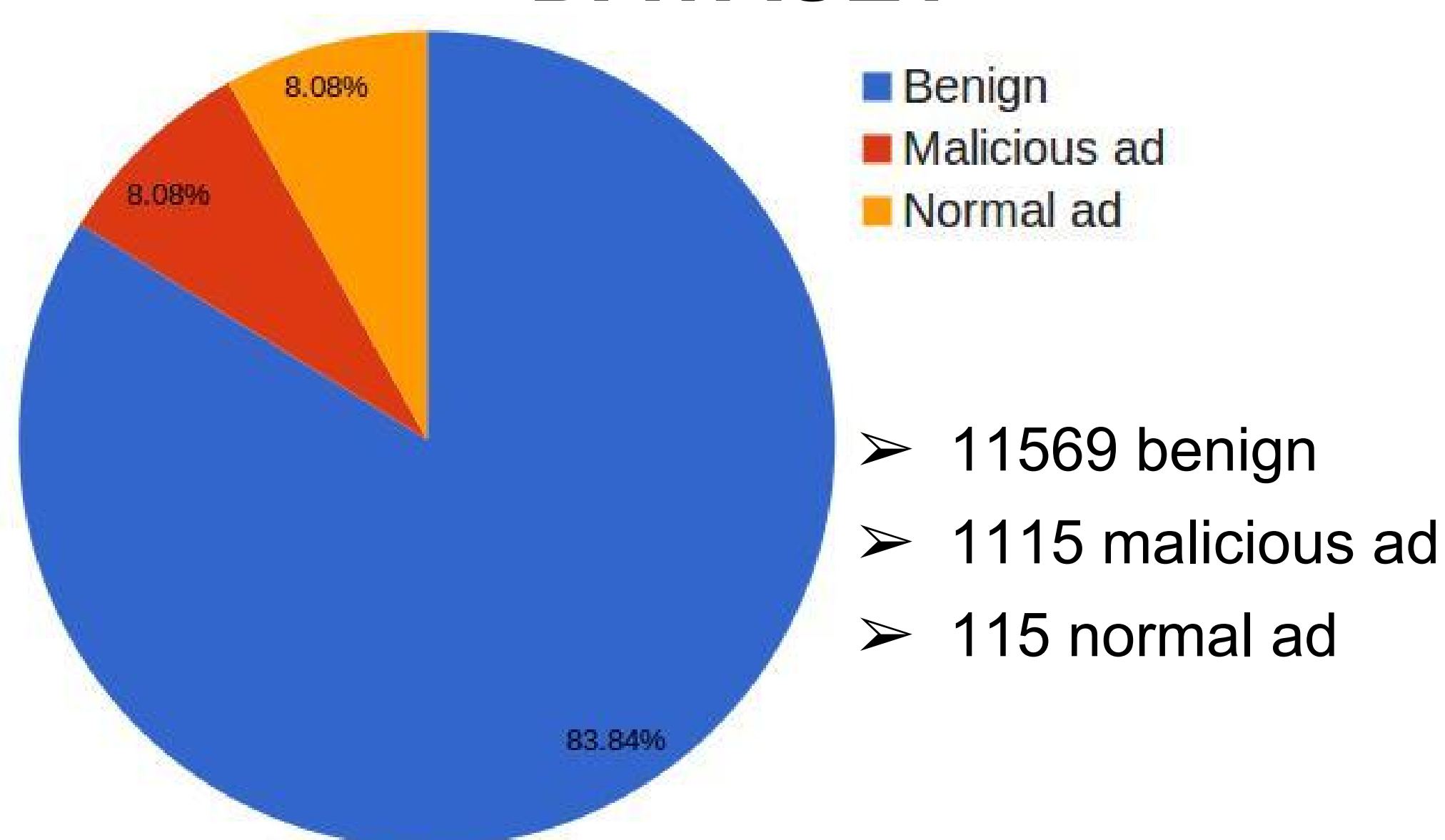
EVALUATION



Classifier	Precision	Recall	Accuracy	FP
RF	0.9977517861	0.9977531347	99.77531347	0.0091301928
SVM	0.9969518964	0.99695586	99.695586	0.0107333154
KNN	0.9847535075	0.9847792998	98.47792998	0.042794723
DT	0.9866146077	0.986591288	98.6591288	0.0344730707
NB	0.8951823224	0.785533087	78.55330869	0.0876893942

Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbour (kNN), Decision Tree J48 (DT), and Naive Bayes (NB).

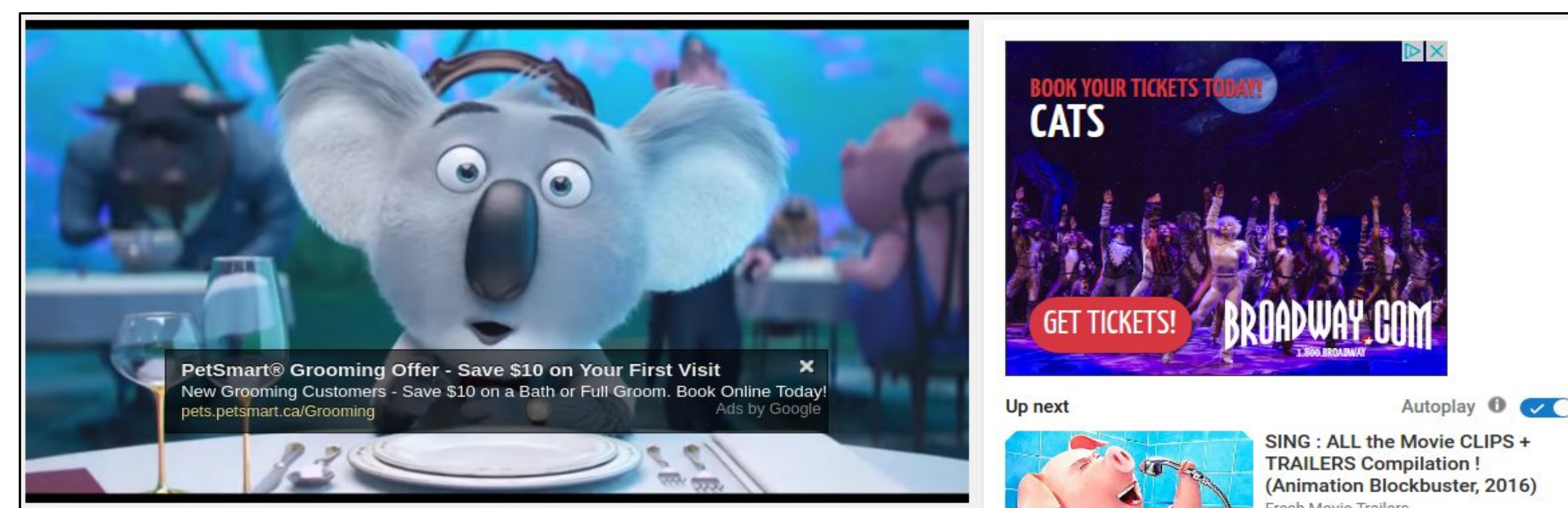
DATASET



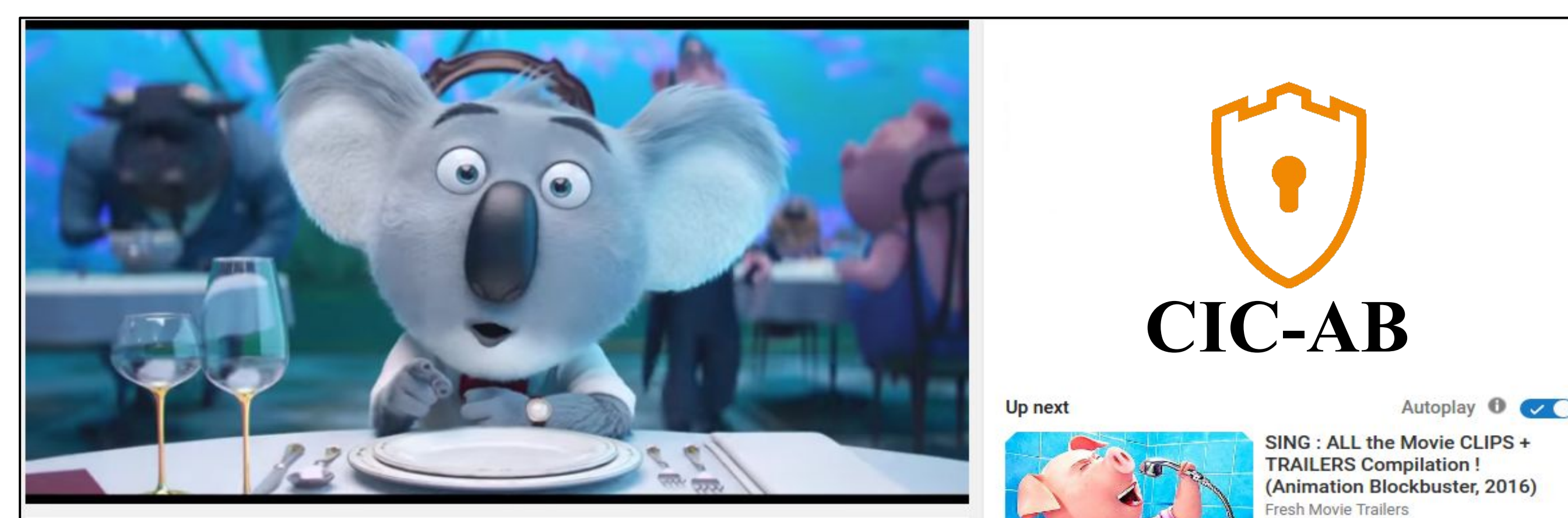
FEATURE SET

Length	No. of special characters	Character continuity
Digit ratio	No. of parameters in query	Alphabet entropy
Length ratio	No. of tokens in domain	Contains a query
Token ratio	No. of digit-letter-digit sequences	Contains 'userinfo' in query
Tokens	No. of letter-digit-letter sequences	No. of digits in query
N-grams	No. of tokens in domain	No. of numeric tokens

IN ACTION



YouTube video without any ad blockers.



YouTube video with CIC-AB.

PROCEDURES

1. Classifiers are built based on the dataset.
2. User visits a webpage.
3. CIC-AB detects an ad in the webpage.
4. CIC-AB blocks or replaces the ad with the CIC-logo.

CONCLUSION AND FUTURE WORKS

- Average accuracy 95%, 3.38% FP rate for five machine learning algorithms.
- CIC-AB in a real-life scenario achieved 90% accuracy in detecting ads correctly.
- We plan to extend CIC-AB to classify ads into different categories: malicious, phishing, and normal(safe).