

Privacy concerns in IoT systems with the use of AI techniques

Fabio Palomba
fpalomba@unisa.it
Università degli studi di Salerno
Salerno, Italy

Giammaria Giordano
ggiordano@unisa.it
Università degli studi di Salerno
Salerno, Italy

Biagio Boi
Gigi Jr Del Monaco
b.boi@studenti.unisa.it
g.delmonaco1@studenti.unisa.it
Università degli studi di Salerno
Salerno, Italy

1 ABSTRACT

The work tries to discover problems related to privacy in IoT systems; in particular those which make use of smart assistances like Alexa or Google Home. The project studies all the existing works in order to understand if these systems have a good level of reliability and integration between each other. A complete refactoring of a system has been proposed; in particular by thinking about the MLOps paradigm, which guarantee a good degree of control.

2 INTRODUCTION

The evolution of smart devices over last years has increased exponentially and the introduction of these devices in the house is progressively growing. The major problem related to these devices is that usually the privacy is not considered, although there are a lot of regulations (just see the GDPR) that describe how the user data have to be stored and who can access to these data. Starting from these two points we've decided to understand what happens within the context of smart assistance. Various projects have been developed and all these projects try to discover a correlation between packets and possible patterns to discover the conversations and the presence of someone inside the home; which can be seen as serious privacy violation.

3 GOAL OF THE PROJECT

The goal of the project is to assess the reliability of developed projects; in particular, an initial possible integration has been evaluated and consequentially; since this integration has not been possible, a comparison between dataset and pipeline automation has been proposed.

4 METHODOLOGICAL STEPS CONDUCTED TO ADDRESS THE GOALS

4.1 First comparison between projects

As introduced, the project started with the comparison among already developed projects and datasets in order to extract important feature from each project. The first considered project has been that one developed by Kennedy et al. [?], which examine a passive attack on home smart speaker able to infer users' voice commands. The project focuses on a particular metric, the so called semantic distance. Anche se questo progetto si è focalizzato su un punto importante della privacy; il ragionamento che sta dietro al calcolo fatto per trovare la distanza semantica sembra essere abbastanza complicato; per questo motivo l'idea di estendere tale progetto è stata trascurata. Il secondo progetto considerato è quello di Alexa real

time analyzer che ci offre una visione più ampia sui dati catturati dalla rete e sulle metriche considerate

5 METHODOLOGY

In order to implement the tool, we've decided to follow each of these steps:

- (1) Consider the current state of art in order to retrieve useful informations. This step is important to produce knowledge to better perform the next steps;
- (2) Analyze the existing datasets to achieve feature engineering;
- (3) Apply normalization techniques (data cleaning, data balancing);
- (4) Implementation and training of a ML model by considering different approaches to find the best fit model for our problem. Looking to the context related projects is most likely that we're going to focus on a Neural Network by using Keras;
- (5) Analyze, monitor and compare the results of each model by using ML Flow tool;
- (6) Develop a real time tool based on this model.

Clearly, all these steps will be conducted using a MLOps approach.