# Continuous behavioral authentication using mouse dynamics based on Artificial Intelligence

**Álvaro Sarmiento Borrajo**

MINISTERIO
DE CIENCIA
E INNOVACIÓN

Financiado por
la Unión Europea
NextGenerationEU

Plan de Recuperación,
Transformación y
Resiliencia

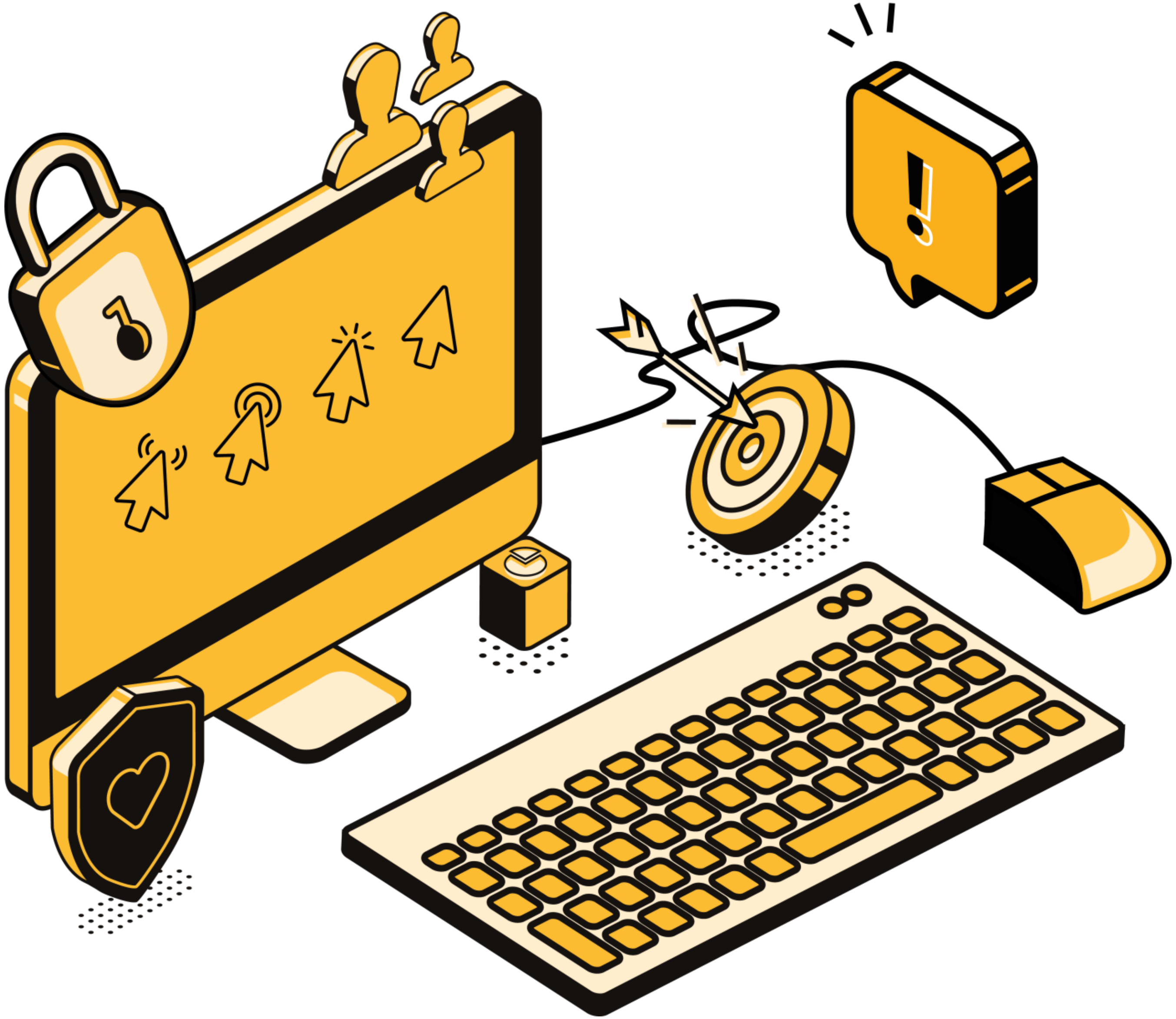
AGENCIA
ESTATAL DE
INVESTIGACIÓN

citic

UNIVERSIDADE DA CORUÑA

# Contents

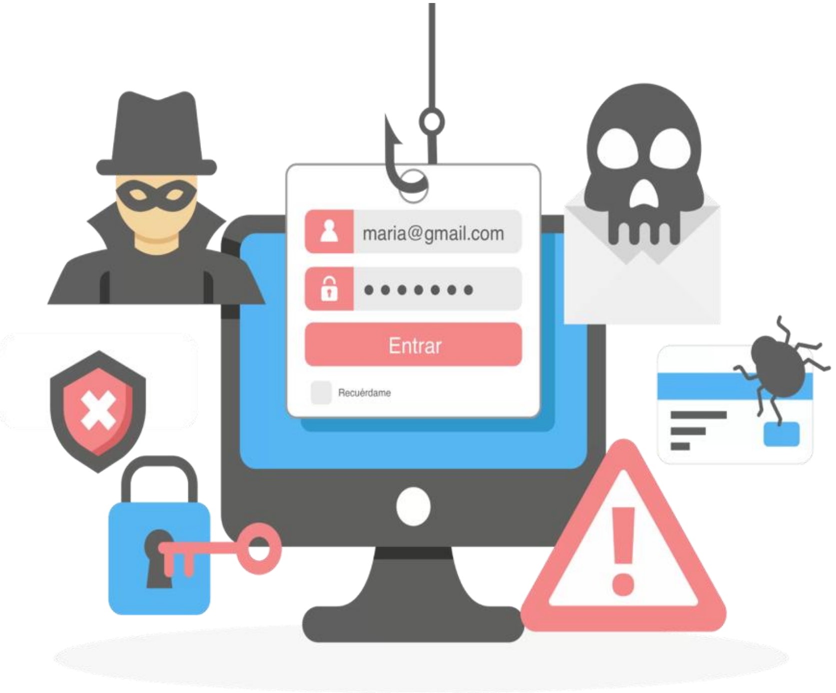

1. Introduction
2. Objectives
3. Architecture
4. Component-based design
5. Data
6. Results
7. Conclusions
8. Future work

# Introduction I

**Authentication systems:**

- Classic schemes:
    - Knowledge-based.
    - Possesion-based.
    - Biometric-based.
- Problem:
    - Identity theft.
- Evolution:
    - Multi-factor approach.
- Persistent problem:
    - Identity theft in open sessions.
- **Solution:**
    - Continuous authentication.



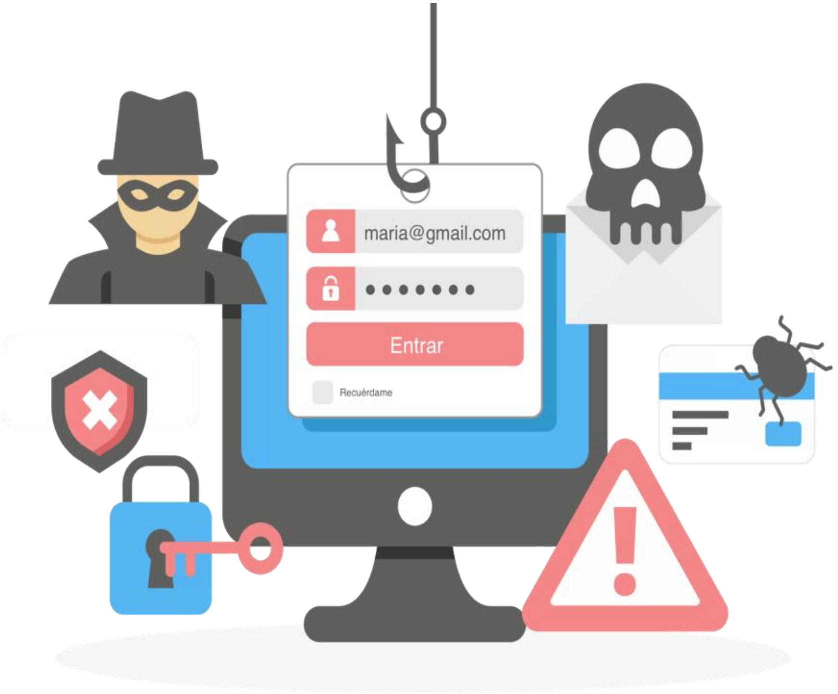
# Introduction II

**Solution:**
Continuous authentication throughout the session.

- Possible approach:
  - Activate the authentication mechanism periodically.
- Problem:
  - Usability: constant interruptions.

- Proposed solution:
  - Use the behaviour of the users themselves.

  - User interactions with input/output devices:
    - Keyboard keystrokes: Too sensitive information!
    - Mouse movement dynamics.



# Main objective

Analysis, design, implementation, and deployment of a continuous authentication platform based on user interaction with the mouse.

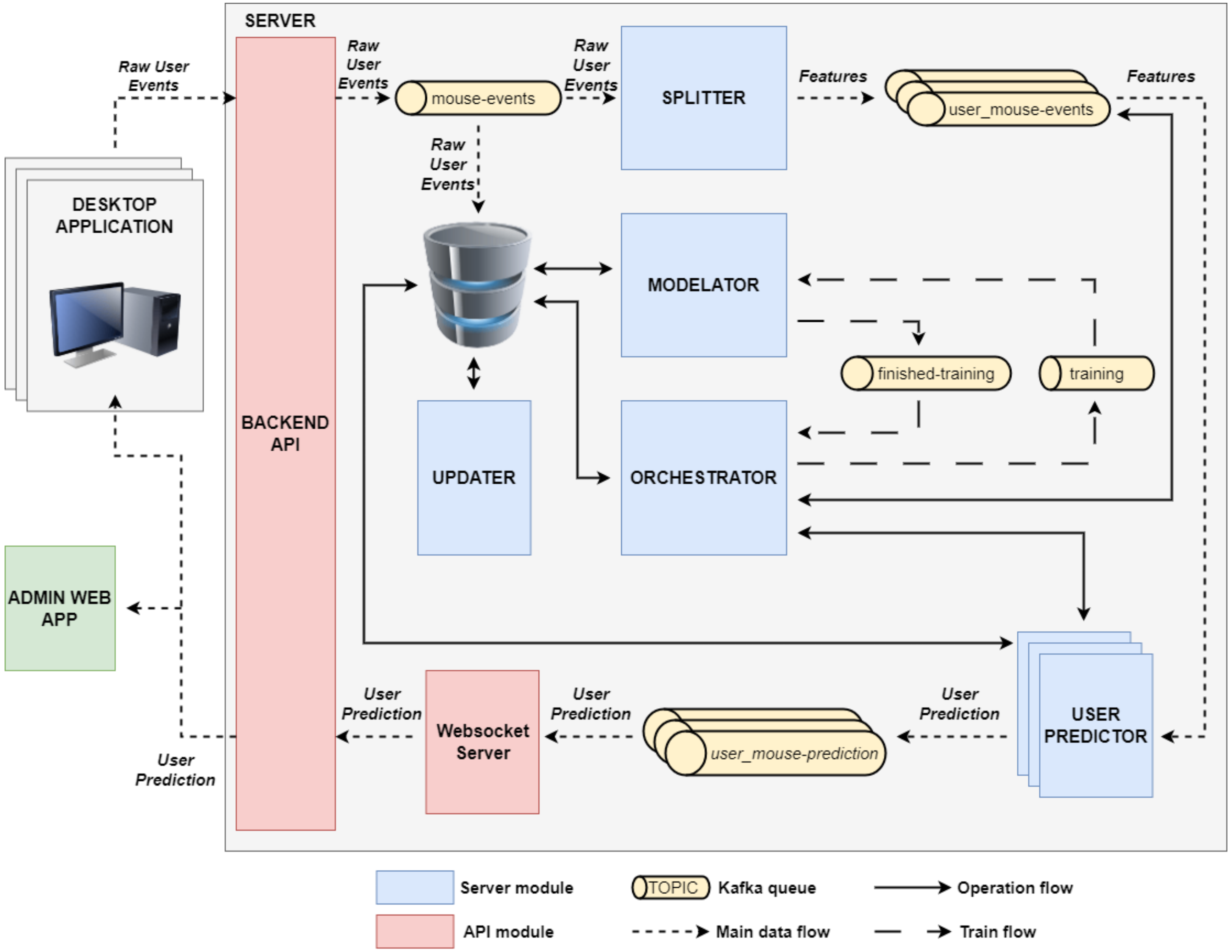
# Sub-objectives

Data extraction and validation

Scalable infrastructure

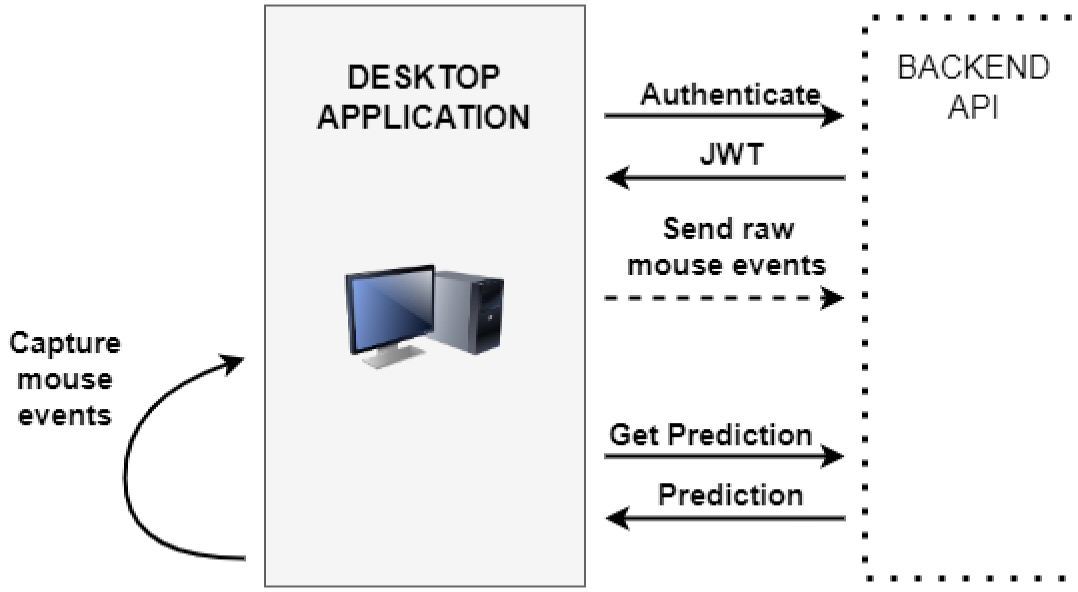
Study of AI models

# Global architecture

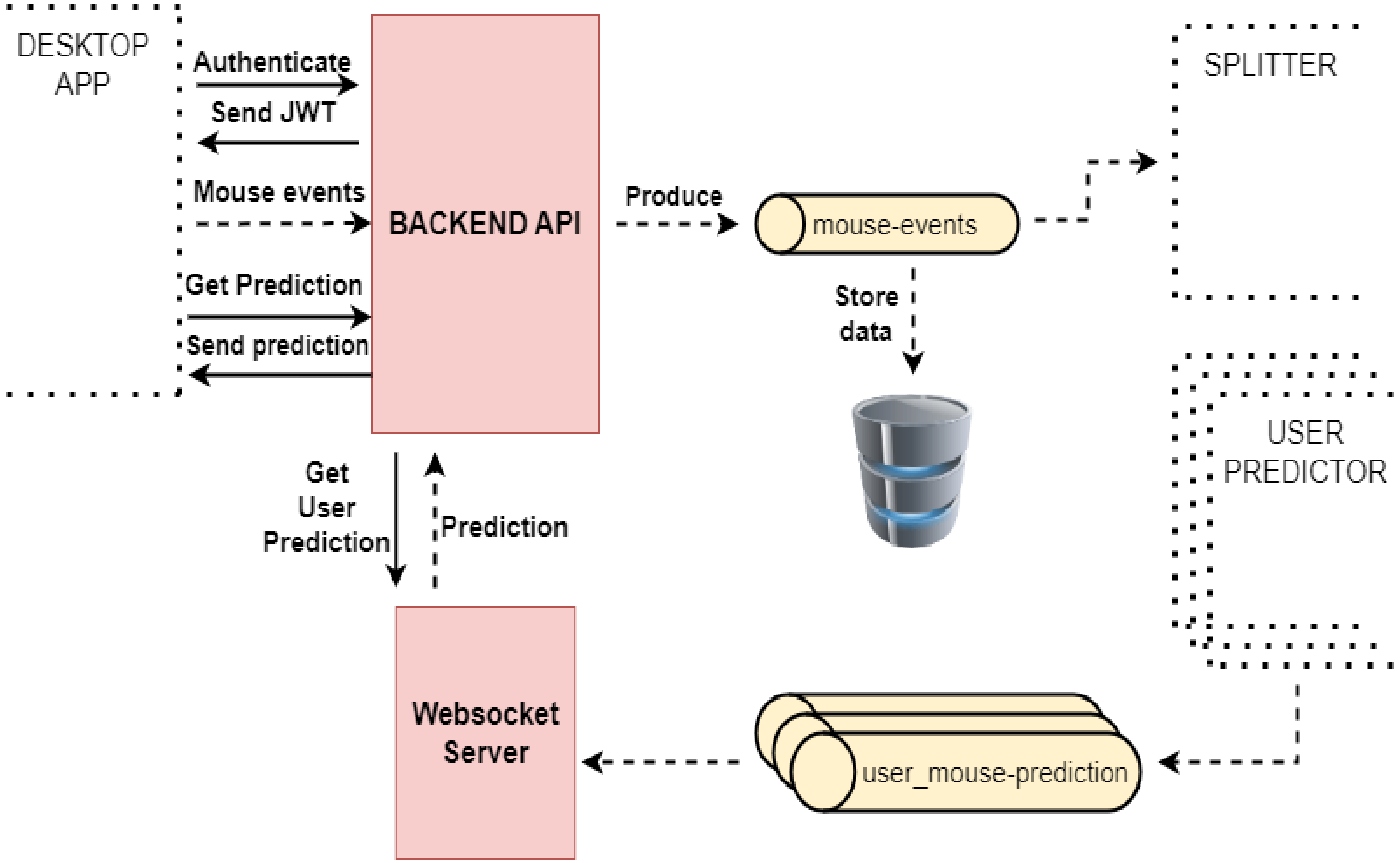


# Component-based design I

- Java implementation.
- Blocks of 1,000 events.
- Prediction: value between 0 and 1.

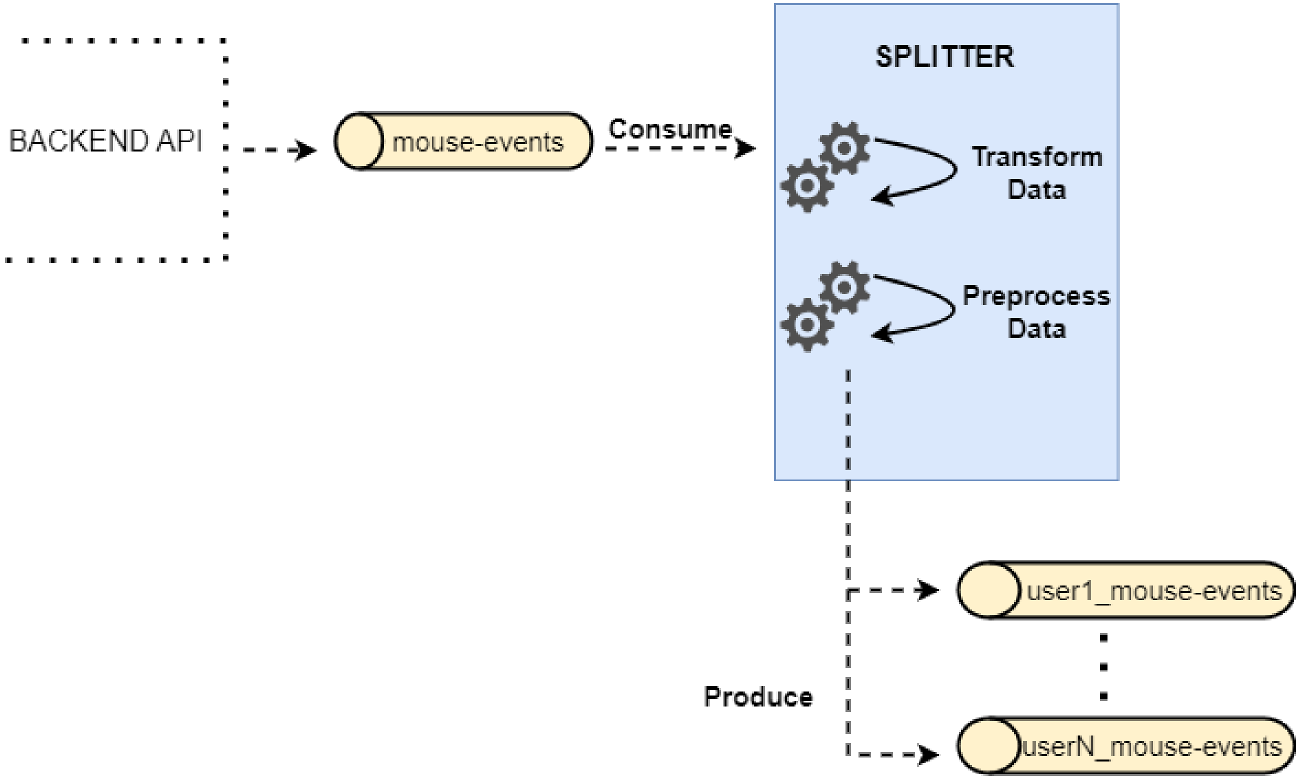


# Component-based design II



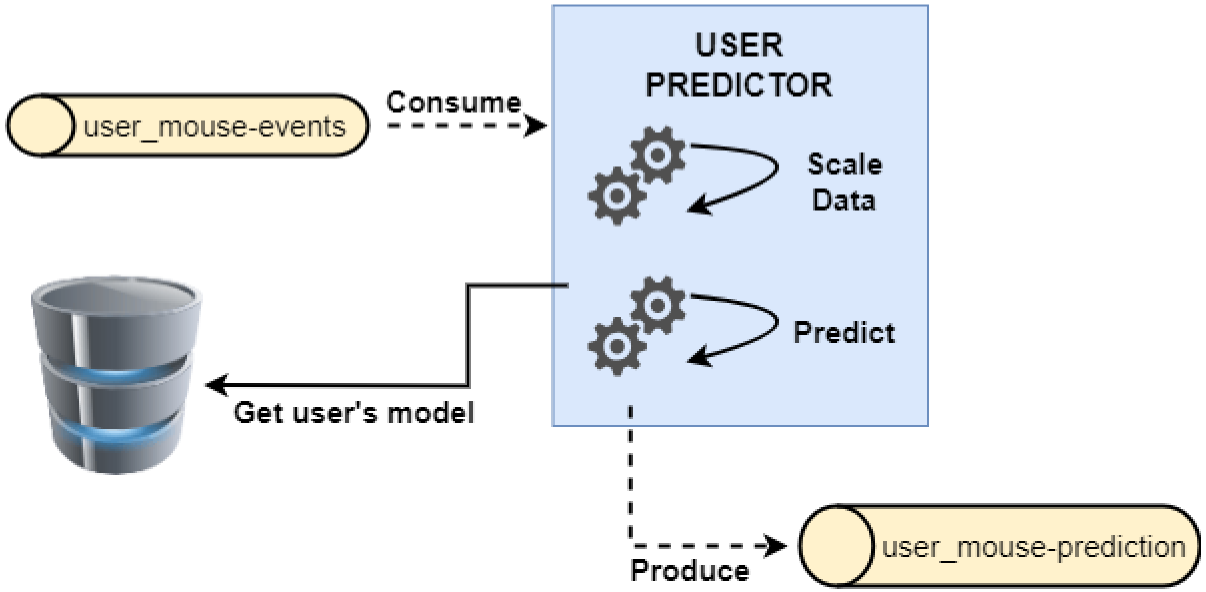
# Component-based design III

- Data preprocessing:
  - Outliers removal.
  - Data clustering.

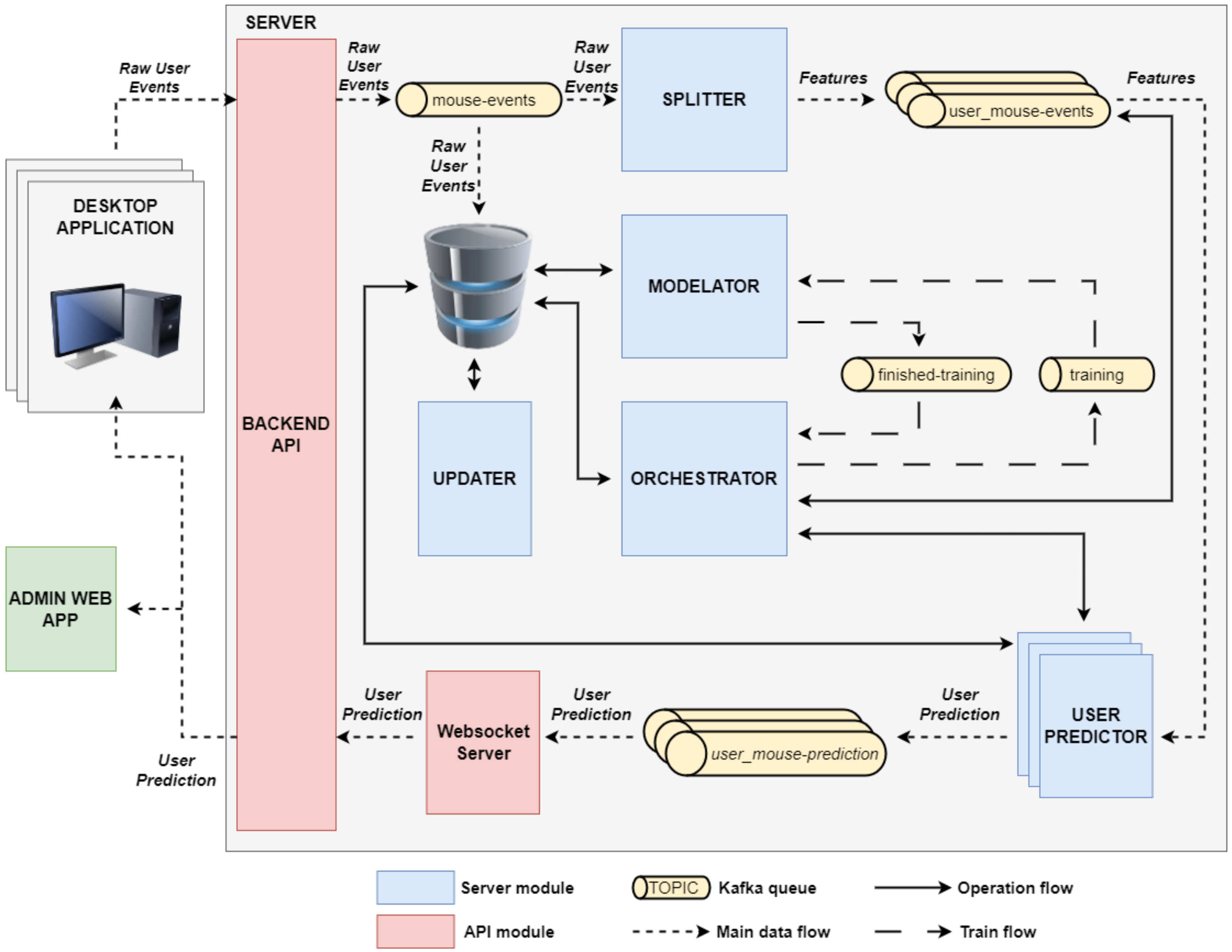


# Component-based design IV

- Data preprocessing:
    - Data scaling.
- **Prediction generation.**

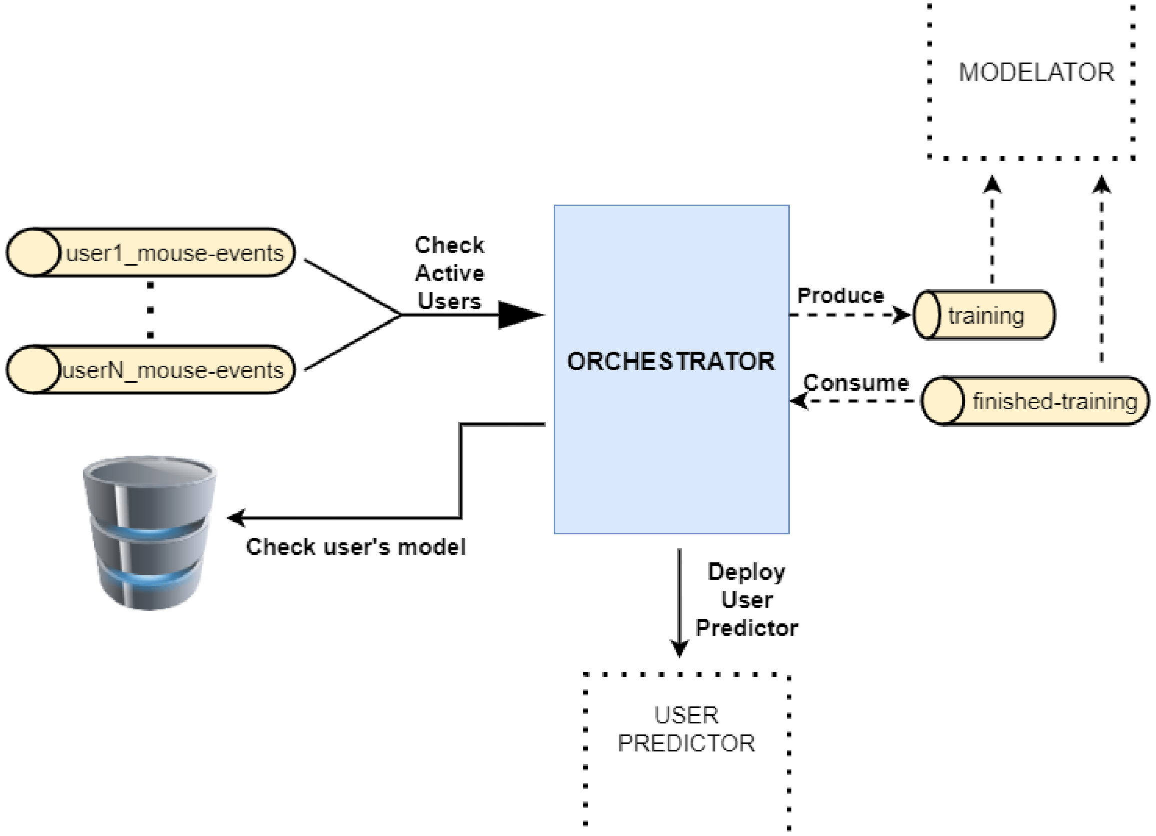


# Global architecture



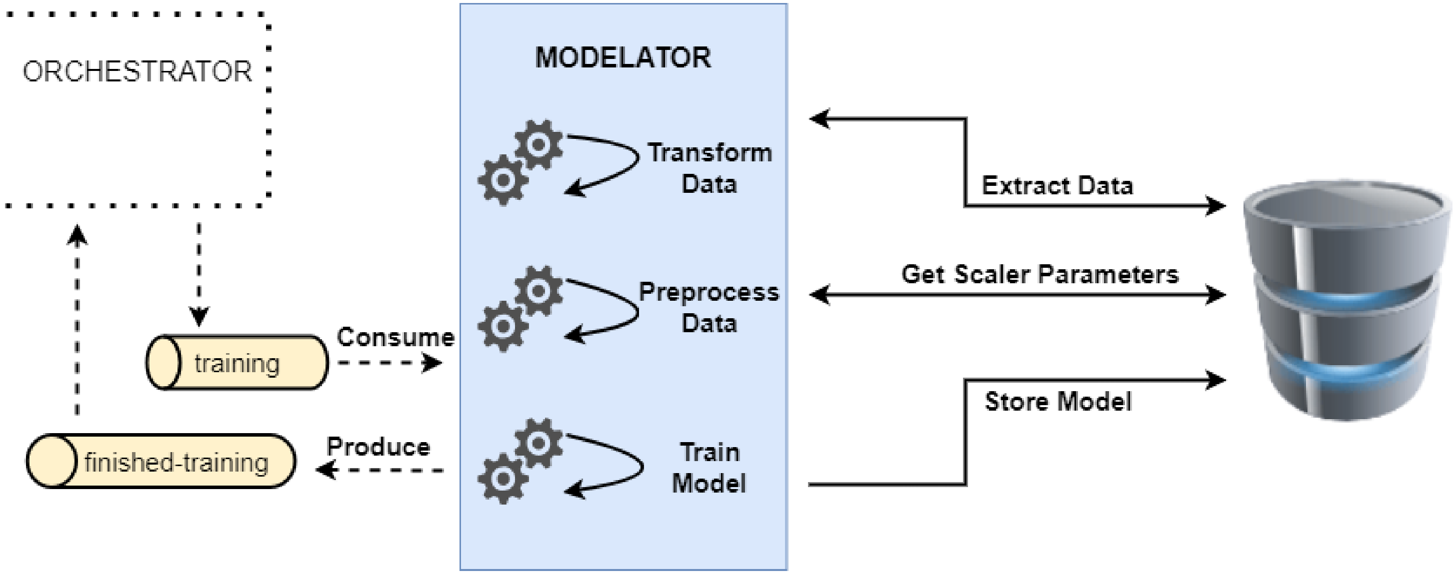
# Component-based design V

- Cases:
  - No model and insufficient data.
  - No model, sufficient data.
  - Model available, but no predictor.



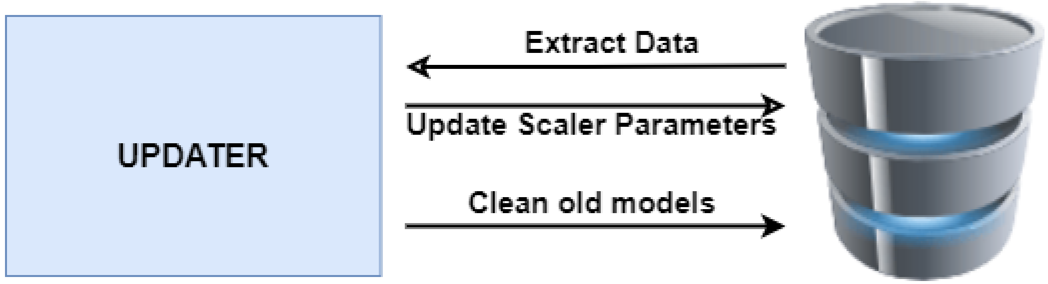
# Component-based design VI

- Resilience to system crashes.



# Component-based design VII

- Update scaling values in short periods.
- Constant but reduced consumption.



# Data I

Event compilation

```
{
    "events": [
        {
            "x": 2256,
            "y": 896,
            "timestamp": 1739910791642
        },
      ...
        {
            "x": 3148,
            "y": 659,
            "timestamp": 1739910821268
        }
    ],
    "metadata": {
        "user": "Ejemplo",
        "mac": "00-00-00-00-00-00",
        "os": "Windows",
        "osVersion": "11",
        "appVersion": "4.0.0.1",
        "timestamp": 1739910821269,
        "UID": "00000000-0000-0000-0000-000000000000"
    }
}
```

# Data II

- Feature extraction.
- Feature processing:
  1. Outliers removal.
  2. Data clustering.
  3. Data scaling.

| Feature | Definition |
|---|---|
| Horizontal velocity | $hv_i = \frac{x_i - x_{i-1}}{t_i - t_{i-1}}$ |
| Vertical velocity | $vv_i = \frac{y_i - y_{i-1}}{t_i - t_{i-1}}$ |
| Tangencial velocity | $tv_i = \sqrt{hv_i^2 + vv_i^2}$ |
| Tangencial acceleration | $ta_i = \frac{tv_i - tv_{i-1}}{t_i - t_{i-1}}$ |
| Tangential jerk | $tj_i = \frac{ta_i - ta_{i-1}}{t_i - t_{i-1}}$ |
| Distance from the origin | $l_i^0 = \sqrt{x_i^2 + y_i^2}$ |
| Distance from the chunk reference | $l_i^{cr} = \sqrt{(x_{cr} - x_i)^2 + (y_{cr} - y_i)^2}$ |
| Slope angle of the tangent from the origin | $\theta_i^0 = \arctan\left(\frac{y_i}{x_i}\right)$ |
| Slope angle of the tangent from the chunk reference | $\theta_i^{cr} = \arctan\left(\frac{y_i - y_{cr}}{x_i - x_{cr}}\right)$ |
| Curvature | $c_i = \frac{\theta_i^0 - \theta_{i-1}^0}{l_i^0 - l_{i-1}^0}$ |
| Curvature rate of change | $\delta c_i = \frac{c_i - c_{i-1}}{l_i^0 - l_{i-1}^0}$ |
| Curvature within the chunk | $c_i^{cr} = \frac{\theta_i^{cr} - \theta_{i-1}^{cr}}{l_i^{cr} - l_{i-1}^{cr}}$ |
| Curvature rate of change within the chunk | $\delta c_i^{cr} = \frac{c_i - c_{i-1}}{l_i^{cr} - l_{i-1}^{cr}}$ |
| Curvature from the origin | $c_i^0 = \frac{\theta_i^0 - \theta_{i-1}^0}{l_i^0}$ |
| Curvature rate of change from the origin | $\delta c_i^0 = \frac{c_i^0 - c_{i-1}^0}{l_i^0}$ |

where $x_i$ and $y_i$ are the mouse positions and $t_i$ is the *timestamp* for a given record.

# Results

Analysis of Artificial Intelligence models

CatBoost

Decision trees

Neural Networks

Random Forest

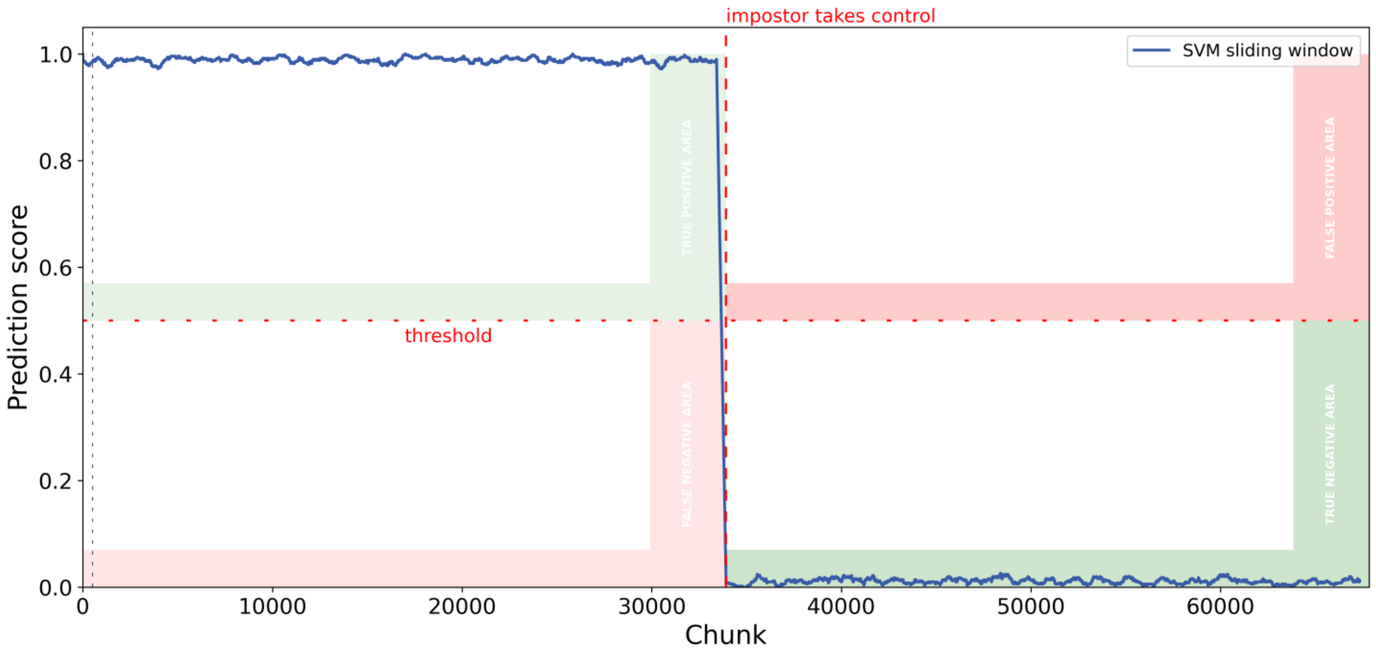
Support Vector Machines (SVM)

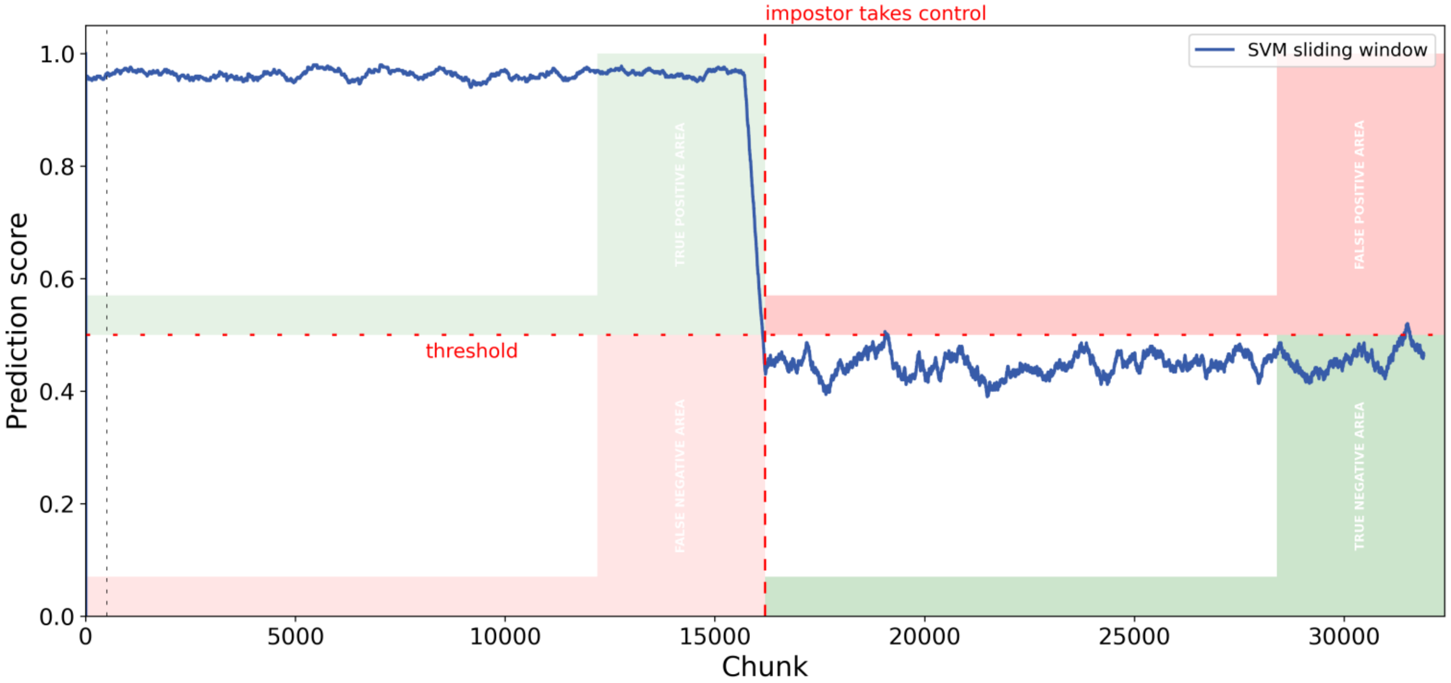
# Results

| | Accuracy | Precision | Recall | F1 | CV Precision (Std Dev) |
|---|---|---|---|---|---|
| CatBoost | 0.9856 | 0.9824 | 0.9894 | 0.9859 | 0.9789 (± 0.0022) |
| Decision Tree | 0.9571 | 0.9618 | 0.9529 | 0.9573 | 0.9562 (± 0.0069) |
| Neural Network | 0.9596 | 0.9524 | 0.9692 | 0.9607 | 0.9488 (± 0.0082) |
| Random Forest | 0.9744 | 0.9703 | 0.9799 | 0.9750 | 0.9644 (± 0.0039) |
| SVM | 0.9790 | 0.9766 | 0.9821 | 0.9794 | 0.9729 (± 0.0033) |

# Results

Rejection Test

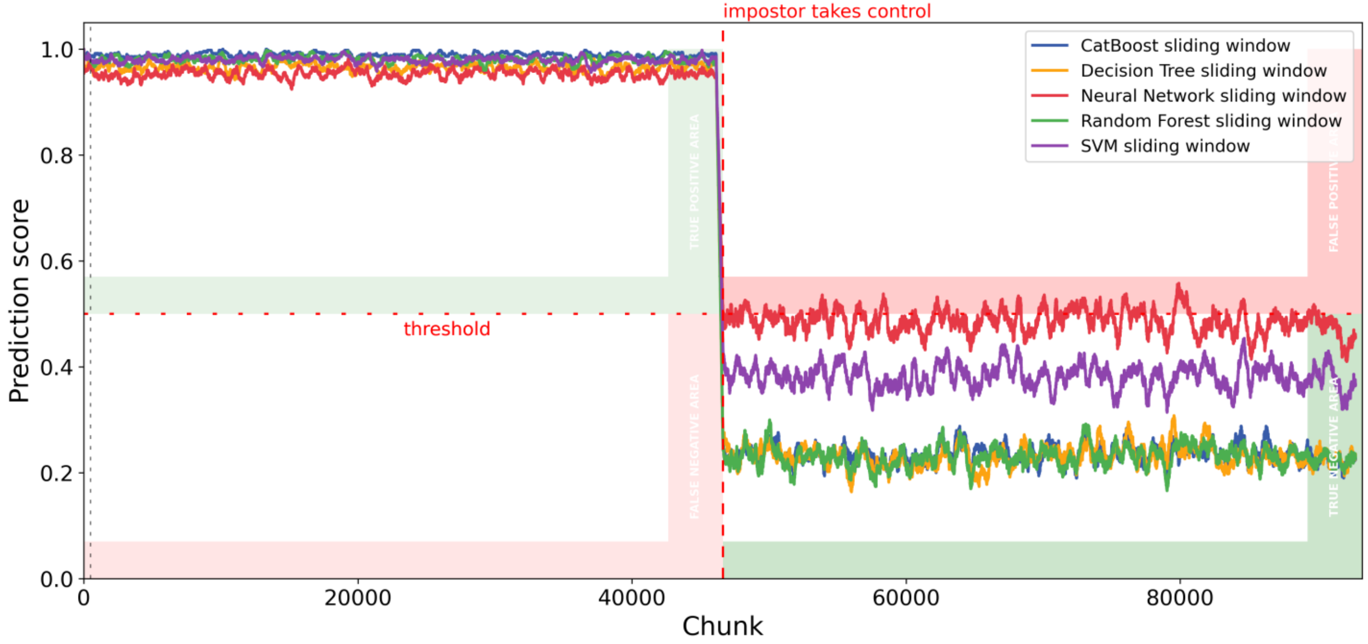


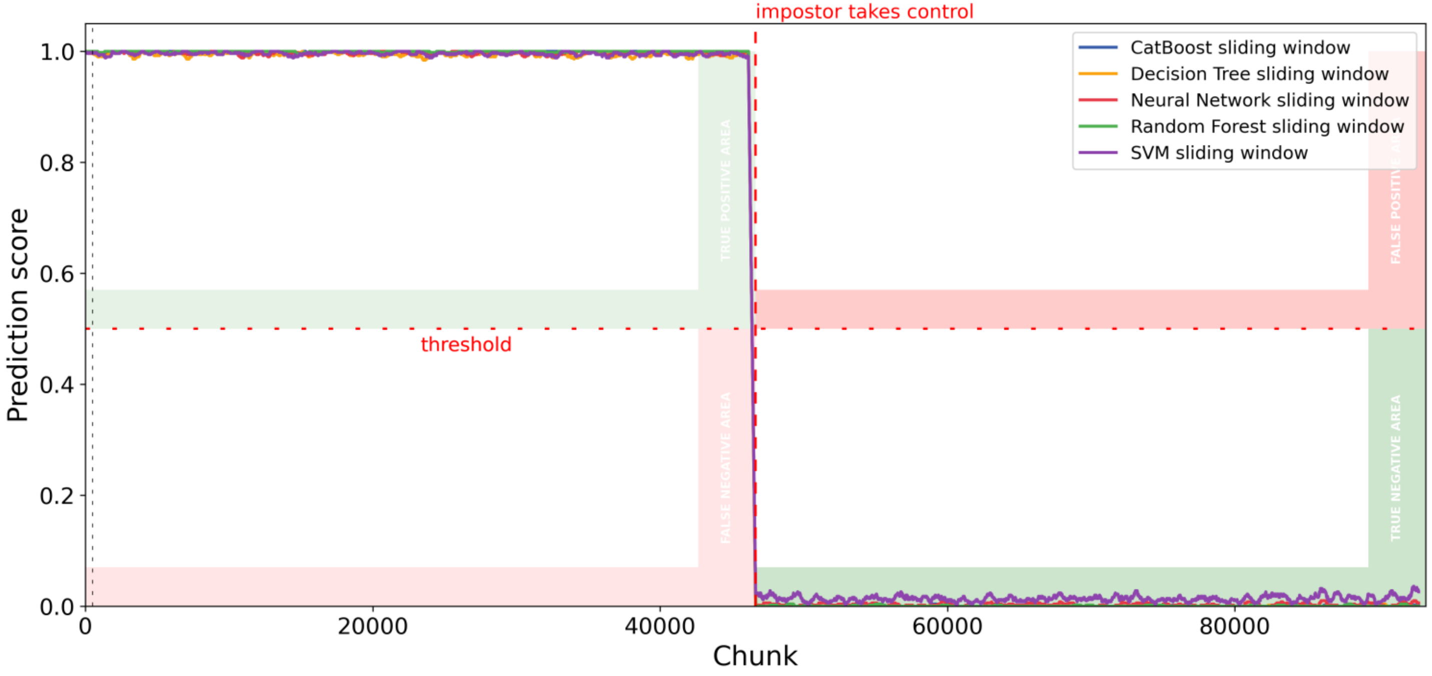



# Results

Rejection Test

Comparison between models



# Results

| User | Organization | App Version | Operating System | Prediction ↑ | Model version | Base Model | Retrain | Last sending of data | Details |
|---|---|---|---|---|---|---|---|---|---|
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 60,52 % | v1 | MLP | | 31 Jul. 2024 13:10:30 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 61,07 % | v4 | MLP | | 31 Jul. 2024 13:11:24 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 11 | 63,52 % | v1 | MLP | | 31 Jul. 2024 13:10:42 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 65,52 % | v4 | MLP | | 31 Jul. 2024 13:11:03 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 67,38 % | v4 | MLP | | 31 Jul. 2024 13:10:58 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 11 | 75,02 % | v1 | MLP | | 31 Jul. 2024 13:11:01 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 75.41 % | v1 | MLP | | 31 Jul. 2024 13:11:07 | |
| [illegible] | [illegible] | 3.2.1.0 | Windows 10 | 79,46 % | v3 | MLP | | 31 Jul. 2024 13:10:41 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 11 | 79,50 % | v1 | MLP | | 31 Jul. 2024 13:10:42 | |
| [illegible] | [illegible] | 3.2.1.2b | Windows 11 | 80,05 % | v1 | MLP | | 31 Jul. 2024 13:11:15 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 80.34 % | v1 | MLP | | 31 Jul. 2024 13:09:42 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 11 | 83,05 % | v1 | MLP | | 31 Jul. 2024 13:11:12 | |
| [illegible] | [illegible] | 3.2.2.0b | Windows 10 | 84.21 % | v4 | MLP | | 31 Jul. 2024 13:10:43 | |

MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
citic
UNIVERSIDADE DA CORUÑA

# Conclusions

- Advances in attacks against traditional authentication systems have underscored the need for more dynamic and secure mechanisms to guarantee user identity, not only at the beginning of the interaction with a system but also throughout active sessions.

- The identity verification through mouse dynamics is feasible and offers strong potential as a complementary mechanism to traditional authentication methods.

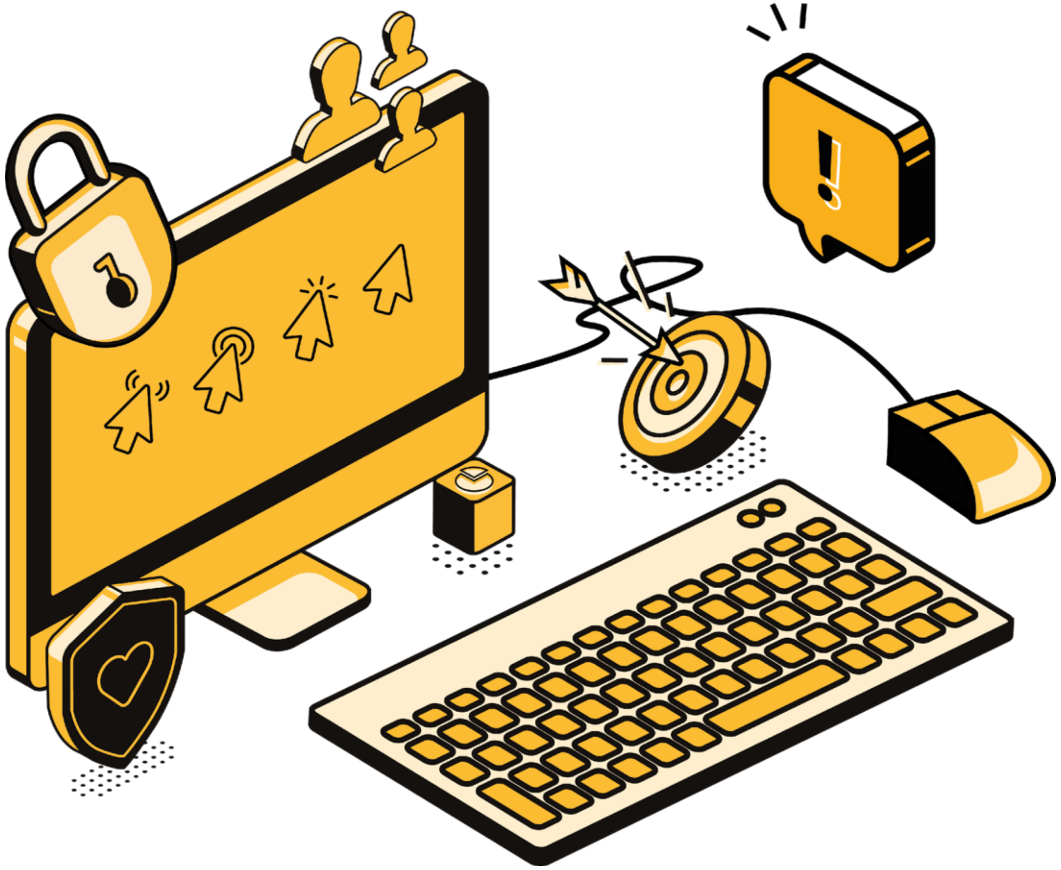
- The results obtained confirm that mouse dynamics can achieve high levels of accuracy, precision, and recall, positioning them as a reliable basis for continuous authentication.

MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
citic
UNIVERSIDADE DA CORUÑA

# Future work

- A systematic performance evaluation of the platform, instrumenting the pipeline to measure how resource consumption evolves as data volume and user concurrency increase.

- Expand the range of behavioral features and validate the platform with larger and more heterogeneous populations.



# Continuous behavioral authentication using mouse dynamics based on Artificial Intelligence

**Álvaro Sarmiento Borrajo**

## Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation MCIN / AEI / 10.13039 / 501100011033, and EUNextGenerationEU / PRTR Ref. TED2021-130492B-C21. We also acknowledge support from the Xunta de Galicia and the European Union (FEDER Galicia 2021-2027 Program) Ref. ED431B 2024/21, ED431B 2024/02, and CITIC ED431G 2023/01.

MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
citic
UNIVERSIDADE DA CORUÑA